Dissecting Functional Heterogeneity in the Human CD8 T Cell Compartment to Characterize Superior Effector Responses

Thesis

Gabriele De Simone

DISSECTING FUNCTIONAL HETEROGENEITY IN THE HUMAN CD8$^+$ T CELL COMPARTMENT TO CHARACTERIZE SUPERIOR EFFECTOR RESPONSES

For the Degree of

Doctor of Philosophy

International PhD Program in Immunology

HUMANITAS CLINICAL AND RESEARCH INSTITUTE

Milan, Italy

Affiliated Research Centre to "The Open University"

Milton Keynes, UK

Under the supervision of

Director of Studies: Dr. Enrico Lugli

External Supervisors: Prof. David A. Price and Dr. Stephen Man

January 2020

Abstract

The constant development of technologies capable to define features at the level of single cells enabled to investigate cellular heterogeneity that was previously inaccessible due to population averaging. In this work, I took advantage of cutting-edge single cell technologies to unravel functional diversification within the human $CD8^+$ T cell compartment, with the final aim to identify T cell subsets that can exert superior effector responses against viruses and cancer. In mice, the capability to exert effector responses correlates with TCR sensitivity for self-Ags, as assessed by indirect measurement of surface expression of CD5. Equivalent functional heterogeneity has not been previously described in humans. We investigated the $CD8^+$ naïve T (T_N) cell pool, and discovered the presence of two discrete subsets with distinct effector differentiation potential, as defined by the differential expression of the chemokine receptor CXCR3. Among these, $CXCR3^+$ (T_NR3^+) cells showed increased cytokine production and effector differentiation potential, both in vitro and in vivo. These features correlated with the expression of T cell receptors suggesting enhanced interactions with p:MHC-I complexes. Thus, functional heterogeneity exists in $CD8^+$ T_N cells, thereby showing that the effector differentiation potential is predetermined at the level of pre-immune repertoire.

Stem-like characteristics of T cells have been a long-standing interest of the lab. T cell memory relies on the generation of antigen-specific precursors with stem-like properties. However, precise definition of these progenitors and how they shape memory and effector differentiation is still unclear. We characterized subsets of clonally, phenotypically, functionally, epigenetically and transcriptionally distinct stem-like $CCR7^+$ $CD8^+$ memory

T cells, identified by the presence or absence of inhibitory receptors PD-1 and TIGIT. We found that PD-1^{-}TIGIT^{-} progenitors were committed to a functional lineage, while PD-1^{+}TIGIT^{+} progenitors were committed to a dysfunctional, exhausted-like lineage. Collectively, these data demonstrate the existence of parallel differentiation programs within the early-differentiated CD8^{+} memory T cell compartment, providing fundamental information for the identification of T cells that are more suitable for cell-based immunotherapies.

Table of contents

1 Introduction

1.1 Adaptive immunity: the T cell compartment

T lymphocytes belong to the adaptive immune system and play a role in cell-mediated immunity. Their progenitors arise from multipotent haematopoietic stem cells (HSC) in the bone marrow (BM) which migrate to the thymus to undergo a process of maturation that includes the rearrangement of the T cell receptor (TCR) genes. TCRs ensure antigen recognition through interactions with peptides processed and presented by major histocompatibility complex (MHC) class II molecules, that are present on antigen-presenting cells (APCs), or class I molecules, that are present on nearly all cells in the body[1]. The TCR is a variable αβ or γδ heterodimer coupled with the non-variable signal transduction CD3 complex, which in turn contains γε, δε, and ζζ subunits[2]. T cells are classified in αβ T lymphocytes, and γδ lymphocytes depending on whether α and β or γ and δ chains compose the heterodimeric TCR[3].

Mature T cells differentiate in the thymus, where different stages of differentiation can be identified on the basis of combinatorial expression of CD4 and CD8 coreceptor molecules. During the double negative phase (DN), $CD4^-$ $CD8^-$ thymocytes rearrange variable (V), diversity (D) and joining (J) segments of the locus encoding TCR β. During the double positive phase (DP), $CD4^+$ $CD8^+$ thymocytes undergo rearrangement of V and J segments of the locus encoding TCRα. The addition and subtraction of nucleotides at the gene segment junctions[4,5] also occurs in order to increase genetic diversity. Rearrangement of these gene segments generates a repertoire of ~2.5×10^8 different TCRs[6]. However, the pool of possible peptides generated by combination of 20 amino acids and

presented by MHC molecules is ideally >10^{15}, and led to hypothesize that high levels of TCR degeneracy are required to cover the full spectrum of possible peptides. In support of this hypothesis, it has been shown that a single TCR is able to recognize >10^6 different peptides[7].

At the DP stage, TCR rearrangement is followed by positive and negative selection of thymocytes[8]. Positive selection is the process by which thymocytes become restricted by the host MHC: only those thymocytes able to bind self-peptide:MHC (p:MHC) complexes with weak but sufficient affinity will survive and will be positively selected[8]. Conversely, T cells bearing TCRs that react too strongly with self-p:MHC complexes will undergo apoptosis, i.e., they will be eliminated by negative selection[8]. This mechanism ensures that only thymocytes capable to recognize self peptides with sufficient, although not too strong affinity can be released in the periphery. It is currently estimated that around 75% of thymocytes are eliminated by negative selection[9]. Mature T cells are thus released in the periphery as naïve T cells (T_N), which retain low levels of reactivity to self p:MHCs. This is insufficient to induce an autoimmune response, but is required to promote their survival and homeostasis by slow proliferation[10,11].

Although pivotal in order to effectively remove potentially autoreactive T cells from the T cell repertoire, thymic negative selection is only 60-70% efficient[12], thus additional peripheral tolerance mechanisms are necessary. Not all self-antigens are expressed at sufficient levels in the thymus to ensure a complete elimination of all autoreactive T cells. Moreover, although negative selection is highly efficient in eliminating T cells of high avidity for self-peptides, low avidity T cells are often spared. A first layer of maintenance of immunological tolerance in the periphery is due to the

abovementioned affinity threshold of TCR-p:MHC interaction. In fact, thymic deletion requires a lower affinity compared to the one required for activation in the periphery[13], thus mature T cells that escape negative selection will show low reactivity to self-ags, and thus be refractory to peripheral activation[14]. The quiescent state of T_N cells, alongside ag ignorance, constitute the intrinsic mechanisms by which peripheral tolerance is exerted prior to TCR engagement. After ag-specific TCR activation, extrinsic mechanisms play an effective role in the maintenance of peripheral tolerance. Self-ags are presented in SLOs to T_N cells by DCs which, in absence of pathogens, are in a quiescent state which is characterized by low levels of costimulatory molecules, such as B7.1 and B7.2, which bind the costimulatory receptor CD28. T cell activation in absence of a second signal (costimulation) results in suboptimal proliferation and effector function capabilities, resulting in the persistence of “anergic” T cells, characterized by a long-term state of hyporesponsiveness and repression of TCR signaling, or elimination by clonal deletion[15].

T_N cells leave the thymus and migrate through the lymph and blood to secondary lymphoid organs (SLO) such as the spleen and lymph nodes, where they continuously scan APCs for the presence of cognate exogeneous antigens[16]. If a cognate antigen is recognized via TCR-p:MHC interaction, T_N cells initiate a process of clonal expansion, characterized by high proliferative rate and the acquisition of effector functions that ensure the elimination of pathogen-infected cells[17]. After antigen is cleared, the vast majority of T_N-derived effector T cells die, and only a small fraction of approximately 5% of these cell will survive, giving rise to a population of clonally expanded long-lived memory T cells

(T_{MEM})[18], that will provide a more rapid recall response compared to T_N and will confer protection against a secondary antigen exposure.

1.2 Heterogeneity and differentiation of the T cell compartment

The T cell memory compartment is characterized by high heterogeneity. Indeed, it is composed by a plethora of subsets that are different in terms of phenotype, functional activity and capability to respond to homeostatic or antigenic simulation[19]. Sallusto and Lanzavecchia proposed that human memory $CD4^+$ and $CD8^+$ T cells can be subdivided into two main subsets, central memory (T_{CM}) and effector memory (T_{EM}), on the basis of their capability to migrate in the body. In fact, T_{CM} express CC chemokine receptor-7 (CCR7) and L-selectin (CD62L), regulating homing towards SLOs, and for this reason are preferentially found at these sites compared to T_{EM}, which instead are more abundant in peripheral tissues. Compared to T_{CM}, T_{EM} cells display increased expression of transcription factors (TFs) and molecules driving terminal differentiation such as Zinc finger E-box-binding homeobox 2 (ZEB2), Inhibitor of DNA-binding 2 (ID2) and PR domain zinc finger protein 1 (*PRDM1*, also known as Blimp-1), among others, display more rapid effector responses, characterized by IFN-γ and TNF but not IL-2 production, and proliferate less extensively[19]. On the other hand, T_{CM} express elevated levels of molecules responsible for long-

term persistence and survival, have increased proliferative potential and are slower in mediating effector functions compared to T_{EM}[20–22].

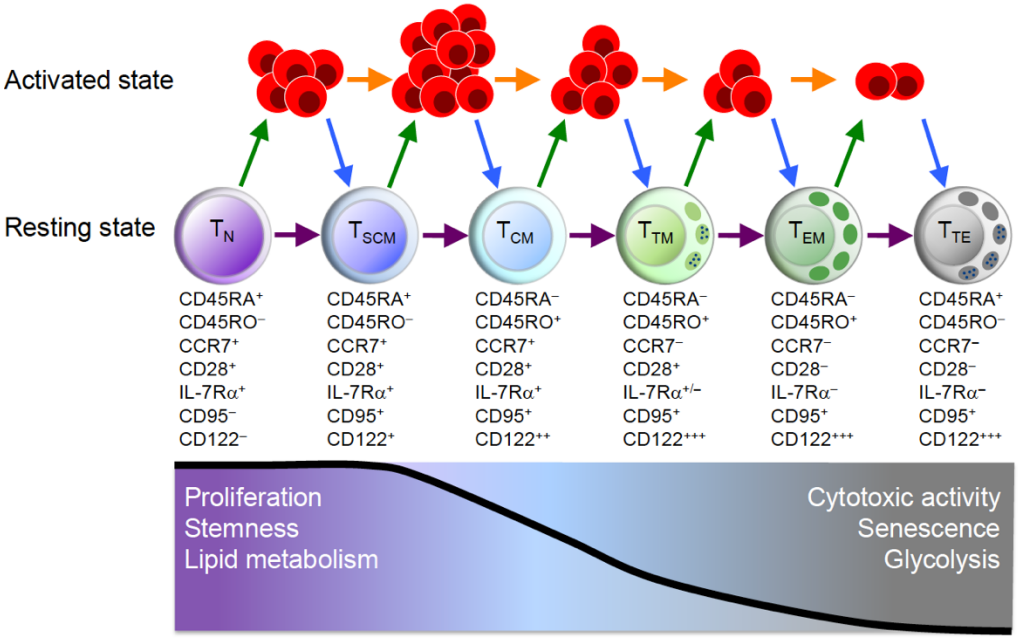


Figure 1: T cell differentiation pattern. Polychromatic flow cytometry analysis allows the identification of each T subset in the T cell compartment. Each T subset is characterized by the expression of specific markers. In red, the markers that change during differentiation from the previous subset. Going from less differentiated T cells to more differentiated ones, functional abilities also change. Adapted from Mahnke et al., Eur J Immunol 2013.

The memory T cell compartment possesses an intrinsic capability of self-maintenance in the long term, according to which T_{CM} were first proposed to behave in a stem cell-like fashion, i.e., to self-renew while generating more differentiated progeny[23]. Recent work in mice, rhesus macaque and humans discovered a new memory T cell population displaying enhanced stem cell-like features compared to T_{CM} and termed T stem cell memory (T_{SCM})[24–26]. This rare subset, representing 3% of all circulating $CD4^+$ and $CD8^+$ T cells in humans, shows a naïve-like phenotype but at the same time possess clearly distinct memory traits such as expression of memory markers CD95 and CD122, lower T Cell Receptor Excision Circles content (TRECs) compared to T_N (indicating higher number of rounds of cell

division) and cytokine producing capacity similar to that of conventional memory T cells. Indeed, upon superantigen stimulation, T_{SCM} display rapid production of IL-2, TNF-α and IFN-γ, however with lower polifunctionality compared to T_{CM} and T_{EM}. T_{SCM} also retain stem cell-like qualities, including a higher capacity to self-renew in vitro and to generate more differentiated progeny compared to T_{CM}. Indeed, after anti-CD3/CD2/CD28 stimulation, which induces T cell differentiation, T_{SCM} demonstrate an increased capacity to maintain their original phenotype when compared with T_{CM}[25], while capable to give rise to all other memory subsets (multipotency)[25,27]. Multiple rounds of antigen stimulation favour the generation of terminal effector T (T_{TE}) cells[27,28], with low proliferation rate, probably as a consequence of extensive telomere shortening[29], among other causes, and poor persistence following transfer in vivo, although still capable to exert potent effector functions upon cognate antigen recognition[27]. Collectively, these observations suggest that T cell differentiation progresses in a linear fashion, according to the relationship $T_N \rightarrow T_{SCM} \rightarrow T_{CM} \rightarrow T_{EM} \rightarrow T_{TE}$ [25,27], and where specific functional capacities are progressively lost, such as proliferative capacity, stemness, and lymphoid homing, or acquired, such as cytotoxic capability, peripheral tissue homing and terminal differentiation/senescence[25,26,30] (Fig 1).

1.3 Homeostasis of human T cells

T cell numbers in the body are tightly regulated by homeostatic proliferation, a turnover occurring at slow rate in the absence of exogenous antigen[31]. In this regard, tonic TCR signalling driven by interactions with self-derived Ags together with additional signals provided by common γ chain (γ_c) cytokines, including IL-2, IL-7 and IL-15, play a

role[10]. While cytokine signalling is sufficient for the homeostasis of memory T cells, both signals are required for the homeostasis of T_N cells. In young individuals, where the thymus is fully functional, the vast majority of peripheral T cells is composed by newly generated T_N cells. With aging, due to thymus atrophy, there is little *de novo* generation of new T cells, thus continuous exposure to environmental Ags causes the accumulation of memory T cells[10].

1.3.1 Naïve T cell homeostasis

The requirement for TCR signalling for T_N cell survival and homeostasis was firstly described more than 20 years ago. Work from Takeda et al demonstrated that newly generated $CD4^+$ T_N cells originated from the thymuses grafted in control immunodeficient RAG $2^{-/-}$ and MHC-$II^{-/-}$ RAG $2^{-/-}$ mice proliferated at the same extent. Nevertheless, $CD4^+$ T cells in the MHC-$II^{-/-}$ RAG $2^{-/-}$ recipient failed to persist, displaying a massive reduction in absolute number over 6 months, thus demonstrating that TCR-p:MHC-II interactions are fundamental tor the maintenance of the T cell pool in the long term[32]. Subsequent work from Tanchot et al identified a similar, although faster decay of murine $CD8^+$ TCR-transgenic T_N cells over time after transfer in MHC-I-deficient hosts, confirming the necessity of TCR-p:MHC interactions for the survival of the T_N cell pool[33].
Alongside TCR signalling, IL-7 plays an essential role for the survival of T_N cells. Several experimental approaches were utilized to demonstrate the fundamental requirement of IL-7 in this regard. Works from Tan et al discovered that both $CD4^+$ and $CD8^+$ T_N cells displayed a lower proliferation rate, and were able to survive only two weeks, according to carboxyfluorescein succinimydil ester (CFSE) dye dilution, after transfer

into syngeneic T-depleted IL-7 deficient mice when compared to those transferred in wild-type (WT) animals[34]. Similarly, injections of anti-IL-7 monoclonal antibodies in thymectomized WT C57BL/6 mice resulted in curtailed cell numbers of T_N cells[35]. On the other hand, transgenic overexpression of IL-7 in mice in vivo revealed an increased number of naïve T cell counts[36]. Collectively, these observations support the notion that IL-7 is necessary for T_N cells survival and that the availability of IL-7 strictly correlates with the extent of the T_N pool.

From a mechanistic point of view, IL-7 binding to CD127 (IL-7Ra) and γ_c induces signals mediated by JAK1 and JAK3 pathways, which in turn lead to the activation of STAT5a and STAT5b, which migrate into the nucleus to regulate transcription[37]. IL-7 is able to promote cell survival by preventing the mitochondrial pathway of apoptosis. BCL-2 and MCL-1 are the main antiapoptotic molecules expressed by T_N cells, and they regulate BAD and BAX, which are apoptotic inducers, capable of triggering the cytochrome-c release from mitochondria and the caspase cascade[38]. IL-7 induces increased expression of BCL-2, as demonstrated by the work of Qin et al using a cutaneous T-cell lymphoma (CTCL) system[39].

1.3.1.1 Homeostatic proliferation of Naïve T cells

As mentioned above, homeostatic proliferation of T_N cells is very slow in physiological conditions (it has been estimated that T_N cells divide every 1 to 6 years[23]), but becomes fast in the context of severe T cell depletion. Such proliferation can be induced by several signals, and thus different types of homeostatic proliferations are recognized[10].

1.3.1.2 Homeostatic proliferation driven by self-p:MHC and IL-7

When the total size of the T cell pool is acutely reduced following irradiation, viral infections or cytotoxic drugs, consumption of IL-7 is drastically decreased, thus leading to increased availability of circulating IL-7[40,41]. In this context, a combination of IL-7 and self-p:MHC elicits increased rates of homeostatic proliferation, according to which T_N cells divide every 24-36 hours[42]. During this Lymphopenia Induced Proliferation (LIP), the proliferating cells acquire phenotypical and functional characteristics of memory T cells[43]. Several studies addressed the importance of p:MHC interactions: according to work from Goldrath and Bevan, polyclonal T cells displayed an impaired capability to undergo LIP upon transfer in TAP-deficient mice expressing low levels of MHC class I, as assessed by CFSE dilution assay[44]. Furthermore, TCR transgenic OT-I T cells underwent LIP even in the absence of cognate Ag, thus highlighting the importance of self-p:MHC interactions for LIP[44].
TCR affinity for self-p:MHC-I ligands also play a role in triggering LIP. According to work from Kieper et al, OT-I T cells, which are characterized by high TCR affinity, were capable to undergo LIP at a fast rate upon transfer in irradiated lymphopenic hosts[45]. On the contrary, T cell lines with lower (2C T cells) or very low TCR affinity (HY cell line) displayed slower rate or almost absent LIP, respectively, when compared to OT-I T cells[45]. Alongside with TCR affinity, several data demonstrated the importance of molecules capable to modulate TCR signalling in the context of LIP. In this regard, T cells lacking the TCR negative regulators CD223 (LAG-3, which binds to high affinity MHC-II), BTLA-4 (which binds to HVEM, a TNF receptor) and SIT (a tyrosine phosphatase 2-binding protein) displayed a faster rate of LIP when compare to WT T cells[46–48]. Costimulatory

molecules are necessary for response to foreign Ags, but they do not play a role in the context of LIP. The only costimulatory molecule that is important for T_N cells in response to self-p:MHC ligands is CD24, since CD24-deficient T cells displayed reduced capacity of homeostatic proliferation in lymphopenic hosts[49].

To summarize, in acute lymphopenic hosts, T_N cells undergo LIP, which relies on self-p:MHC interactions of relatively high affinity, and on increased concentrations and availability of IL-7, in the absence of costimulatory signals.

1.3.1.3 Homeostatic proliferation driven by other cytokines

Besides IL-7, T_N cells can undergo homeostatic proliferation also in the presence of high concentrations of other γ_c cytokines such as IL-2 and IL-15, even in the absence of lymphopenia. IL-2 and IL-15 share two of their three receptor chains, the IL-2/15 Rβ (CD122) and the common γ_c (CD132)[50]. IL-2 and IL-15 also share downstream signalling, which involves the induction of mitogen-activated protein kinase (MAPK) and BCL-2, and the phosphorylation of spleen tyrosine kinase (Syk) and of lymphocyte-activated protein tyrosine kinase (Lck)[50]. The most prevalent form of IL-15 in vivo is a membrane-bound form, where IL-15 binds with high affinity to IL-15Rα, thereby signalling to target cells by trans-presentation[51]. In the non-complexed form, however, IL-15 can still bind the 15Rβγ signalling complex, although at lower affinity. Receptor binding triggers the activation of PI3K and MAPK signalling pathways[52].

A role for IL-2 and IL-15 in sustaining T cells homeostatic proliferation were first discovered more than 10 years ago. Work from Cho et al, demonstrated that small amounts (10^6) of WT T_N cells injected in non-

irradiated CD122$^{-/-}$ mice proliferated massively, around 100-fold more when compared to acutely lymphopenic (irradiated) B6 hosts[53]. This fast tempo of proliferation was significantly higher in CD8^{+} than in CD4^{+} T cells, and required self-p:MHC interactions[53]. Interestingly, in this scenario of Cytokine Induced Proliferation (CIP), cells proliferated at a rate which is comparable to that occurring with foreign Ag stimulation, and the resulting progeny rapidly acquire a T_{EM} or T_{CM} phenotypes in the presence of high concentrations of IL-2 and IL-15, respectively[53,54]. Notably, *Il-2ra*$^{-/-}$ and *Il-2rb*$^{-/-}$-deficient mice display massive lymphadenopathy, thus indicating that this kind of homeostatic proliferation does not strictly require lymphopenic conditions, and that the quality of the proliferative response is a reflection of the stimulatory capacity of the cytokines involved[10]. Regarding this, IL-2 and IL-15 represent stronger proliferative and differentiative signals compared to IL-7, thus their production needs to be strictly regulated in order to avoid massive activation of bystander T_N cells in the context of strong immune responses[10].

Under physiological conditions, IL-2 has a minor role in controlling T cell homeostasis, except for regulatory T cells (Treg), while IL-15 promotes T_N homeostasis, as assessed by work from Berard et al which demonstrated that in IL-15-deficient mice total count if T_N cells was highly impaired[55].

1.3.1.4 Homeostatic proliferation driven by commensal Ags

As previously stated, homeostatic proliferation relies mainly on self-p:MHC interactions[10]. Nevertheless, also foreign Ags play a role in this regard[56]. In particular, when WT mice are rendered acutely lymphopenic by irradiation, LIP is triggered by self rather than foreign Ags[10]. On the contrary, foreign Ags drive a chronic Lymphopenia-Induced Proliferation

(cLIP) in chronically lymphopenic hosts, such as RAG-deficient, SCID, nude mice or TCR deficient mice[10]. This was supported by the evidence that that cLIP, but not LIP is abolished in mice raised in germ-free conditions, thus suggesting that the source of Ags is provided by the commensal microflora of the chronically lymphopenic hosts[56]. Additional evidence that cLIP is triggered by foreign Ags is provided by the fact that proliferating cells required CD28 costimulation and rapidly upregulated activation markers, such as CD25, and acquired effector functions, according to TNF-α and IFN-γ production[56].

1.3.2 Homeostasis of memory T cells

Numbers of memory T cells tend to increase with age, as a consequence of the response of T_N cells to both foreign and self-Ags. Both $CD8^+$ and $CD4^+$ memory T cells undergo intermittent cell divisions once every 2-3 weeks, which is paralleled by a balancing degree of cell death[57]. This turnover is largely MHC independent and relies on a combination of contacts with IL-7 and IL-15. First demonstrations regarding the factors involved in the control of memory T cells homeostasis date back to 1990s, where in vivo activation of the immune system caused by viral infections or injection of LPS or poly I:C without Ags reflected in a massive increase in the proliferation rate of both Ag-specific and bystander memory T cells[58–60]. Specifically, production of Type I IFN increased after adjuvant stimulation, thus increasing IL-15 production by APCs, which in turn can directly stimulate memory T cells to elicit their proliferation[59]. As mentioned earlier, IL-15 (and IL-2) binds to CD122 and CD132, that are highly expressed on $CD8^+$ memory T cells, less so on $CD4^+$ memory T cells[60]. Work from Intlekofer et al demonstrated that CD122 expression is

regulated by T-bet and Eomes, two transcription factors that are upregulated upon T cell activation and differentiation[61]. In this work, *Eomes*$^{+/-}$*T-bet*$^{-/-}$ mice displayed massive depletion of IL-15-dependent cell subsets, such as CD8^{+} memory T and NK cells, similarly to what occurs in IL-15-deficient mice, thereby linking the activity of transcription factors important for CD8^{+} effector T cell differentiation to survival[61]. IL-15 is extremely important for the survival of Ag-specific memory T cells. According to work from Ahmed and colleagues, a population of functional Ag-specific memory T cells is capable to develop in IL-15-deficient mice after lymphocytic choriomeningitis virus (LCMV) infection, but its total numbers was drastically reduced and then disappeared over time, thus showing that IL-15 is dispensable for the generation of Ag-specific memory T cells, but is necessary for their long term homeostatic proliferation[62].

Homeostasis of memory T cells relies on IL-7 as well as on IL-15. According to Kieper et al, overexpression of IL-7 in IL-15-deficient mice restored normal numbers of memory T cells, whereas background levels of IL-7 were not sufficient to sustain their generation and survival[36]. Work from Geginat and colleagues investigated how acquisition of several functions and phenotypical features of human CD8^{+} T cell subsets is affected by cytokine stimulation in vitro[51]. The authors showed that proliferative response to IL-7 was low in all subsets analysed, while response to IL-15 increased in parallel with progressive differentiation, as assessed by CFSE dilution. Further addition of IL-2 selectively stimulated memory T cells and provided additional boost to IL-7 and IL-15 responses[51], indicating that γ_c cytokines are not redundant, rather can act in concert to support effector T cell functions.

Homeostasis of memory $CD4^+$ T cells is broadly similar to memory $CD8^+$ T cells since they both need a combination of IL-7 and IL-15. However, there is a difference in terms of IL-7 and IL-15 usage: in fact, $CD4^+$ memory T cells display lower expression of CD122, thus they are less IL-15-dependent and their basal homeostatic proliferation requires both IL-7 and IL-15[63].

1.4 Heterogeneity of human T_N cells

T_N cells have long been considered a largely homogeneous population[64–66]. Nevertheless, the recent development of technological approaches capable to measure features at the single-cell level revealed heterogeneity in terms of response to antigen stimulation by individual T_N cells. Work from Buchholz et al[67] and Gerlach et al[68] revealed that, in single-cell adoptive transfer or barcoding experiments in the context of mouse models of infection, individual $CD8^+$ T_N cells displayed different proliferation patterns. While some clones displayed massive proliferation, others proliferated at a lesser extent, and the entity of this proliferation positively correlated with effector versus memory differentiation[67,68]. Thus, single T_N cells can behave differently in response to the same infectious stimulus.

A different study from Tubo and colleagues obtained similar results for murine $CD4^+$ T cells. Furthermore, this work demonstrated that TCR-p:MHC-II dwell time or the antigen dose may play a role in differentiation divergence[69].

Additional studies on the murine T_N cell repertoire revealed that the ability to respond to foreign antigens can be predicted by the level of cross-reactivity to self-antigens. In both $CD4^+$ and $CD8^+$, increased T_N cell

affinity for self Ags correlated with increased surface expression of CD5[70–72]. Accordingly, low and high expression of CD5 identified two T_N cell populations, where $CD5^{hi}$ $CD8^+$ T_N are hyper-responsive to IL-2 and IL-7 homeostatic cytokines [73] and have increased levels of genes involved in effector differentiation[72] compared to $CD5^{lo}$ cells. In $CD4^+$, $CD5^{hi}$ T_N displayed enhanced TCR-dependent signaling potency[70,71]. Moreover, work from Alanio and colleagues proposed CD5 as a functional analogue for immune phenotypization of human T_N cells in the context of chronic infection[74]. Indeed, they identified lower CD5 expression on $CD8^+$ T_N cells from patients with chronic HCV infection compared to healthy controls, and they correlated such reduced expression with a lower threshold for TCR signaling, which in turn brought to the accumulation of antigen-inexperienced memory-phenotype T cells in this cohort, as proposed by the authors[74]. However, it is currently unclear whether CD5 marks distinct subsets of T_N cells also in humans.

1.5 T cell exhaustion and heterogeneity of exhausted T cells

Persistent antigen stimulation via the TCR, as occurring in the context or chronic viral infections and cancer, induces a dysfunctional state, generally referred to as T cell exhaustion[75], characterized by the acquisition of distinct gene expression and epigenetic programs, reduced effector capacity and proliferation, and augmented expression of inhibitory receptors including programmed cell death protein 1 (PD-1/CD279), lymphocyte-activation gene 3 (LAG3), T cell immunoreceptor with Ig and ITIM domains (TIGIT), 2B4/CD244, T cell immunoglobulin and mucin domain-containing protein 3 (TIM3) and cytotoxic T-lymphocyte antigen 4

(CTLA4)[76]. In the context of chronic viral infections, virus-specific T cells become progressively exhausted due to persistent Ag stimulation[77]. It is important to stress that these exhausted T cells (T_{EX}) are not fully unresponsive as they are still capable to exert some effector functions, which are thought to be important to contain viral replication while reducing immune activation and thus immunopathology. In this regard, Frebel et al described fatal CD8 T cell-mediated immunopathology early after LCMV infection in mice lacking PD-1 signalling[78]. Thus, T cell exhaustion cannot be considered a fully dysfunctional state, rather a hypo-responsive cell state, still capable to control pathology[79].

It has been recently proposed that T_{EX} cells in chronic infections and cancer are characterised by substantial heterogeneity. Accordingly, Paley et al originally demonstrated that virus-specific $CD8^+$ T_{EX} are composed by a $Tbet^{hi}PD\text{-}1^{int}$ progenitor which differentiates in an $EOMES^{hi}PD\text{-}1^{hi}$ terminally differentiated, non-proliferating progeny[80]. Importantly, depletion of either subset resulted in the failure to control viral replication, thereby showing that both subsets play a central role[80]. Subsequently, several works investigated T-cell heterogeneity during chronic infections and described transcription factor TCF1 as a pivotal regulator of the generation of progenitors[81–86]. According to work from Ahmed and colleagues, $TCF1^+$ $CD8^+$ T, identified also as $PD\text{-}1^+CXCR5^+TIM3^-$, displayed a memory/stem cell-like phenotype and possessed capability to self-renew and to give rise to a $TCF1^{lo/neg}$ terminally differentiated progeny characterized by a $PD\text{-}1^+CXCR5^-TIM3^+$ phenotype[82]. These $TCF1^+$ progenitors resulted capable to proliferate in response to inhibitory receptors blockade (PD-1/PD-L1) while $TCF1^-$ failed to do so, thus demonstrating that this memory/stem

cell-like population of progenitors constitutes the main effective target of immunotherapeutic treatments in chronic infections[82].

Additional studies in the context of cancer demonstrated that $CD8^+$ tumor-infiltrating T lymphocytes (TILs) possess exhaustion traits common to T cells in chronic viral infections, such as impaired production of effector cytokines and cytotoxic molecules, and increased expression of inhibitory receptors[87]. Recent work from our group identified T_{EX} progenitors within the $CD8^+$ infiltrate from non-small cell lung cancer patients, defined as $CXCR5^+TIM3^-$. Despite partially exhausted, as revealed by the overexpression of several inhibitory receptor molecules, these cells also expressed TCF1 and displayed memory-like hallmarks, such as IL-2 production and enhanced self-renewal capacity compared to terminally-differentiated $CXCR5^-TIM3^+$ cells[88].

Other recent studies identified $TCF1^+$ TILs with memory/stem-like features and cytotoxic potential in human tumours[84,89–91]. In particular, Li et al combined TCR-sequencing with single-cell RNA-sequencing to demonstrate that $TCF1^+$ TILs included bystander cytotoxic populations, while T cell clones specific for tumour Ags displayed a $TCF1^{lo}PD\text{-}1^+LAG3^+$ dysfunctional phenotype, and are characterized by CD39 expression, thus unveiling an additional layer of heterogeneity among human TILs[92].

$CD8^+$ T cell subsets in human are characterized by specific epigenetic programs that dictate transcriptional, functional and phenotypical regulation[93]. Coherently, T_{EX} cells displayed distinct patterns of chromatin accessibility compared to T_N, T_{EFF} and T_{MEM} $CD8^+$ T cells, characterized by the presence of differentially accessible regions in genes associated to inhibitory receptors and transcription factors associated to exhaustion[89,94–98].

Among transcription factors, recent work identified TOX as a central driver of the exhaustion program[99–102] and found that TOX-deficiency in $CD8^+$ T cells improves tumor control[100,103]. Overall, these studies support the notion that exhaustion in chronic infection and in cancer constitutes a separate branch of T cell differentiation with its own progenitors and terminally differentiated cells and characterized by a unique transcriptional and epigenetic identity. Whether a pre-commitment to this exhaustion program can be identified under physiological conditions is still unknown.

2 Aim of the study

The $CD8^+$ T cell compartment is highly heterogeneous in terms of phenotypes and functional capabilities, and the recent increased utilization of single cell techniques allowed a further investigation of such heterogeneity.

In the murine $CD8^+$ T_N pool, CD5 allows the identification of functionally distinct subsets of T_N cells with enhanced capability to respond to foreign antigens, suggesting that predetermined heterogeneity dictates the capacity and the magnitude of the immune response. Whether such functional heterogeneity is also present in the human $CD8^+$ counterpart and can shape immune responses is still unknown. In the first part of this project, we evaluated the functional heterogeneity of the human $CD8^+$ T_N pool of healthy individuals. Specifically, we used flow cytometry and cell sorting for the characterization and isolation of discrete subsets of human $CD8^+$ T_N subsets, as well as molecular technologies, functional assays (both in vitro and in vivo) and RNA sequencing to obtain an all-encompassing characterization of the identified populations.

In the second part of the work, we aimed to characterize the stem-like pool of $CD8^+$ T cells in light of the recent discoveries of stem-like progenitors with distinct traits in cancer and chronic viral infections. We hypothesized that differentiation of these progenitors could occur in physiology in healthy individuals. We applied un unbiased approach guided by single-cell RNA sequencing to identify candidate memory progenitor subsets. We then performed cellular and molecular assays to characterize such progenitors at the functional, transcriptional and epigenetic level, in order to dissect their functional heterogeneity and their fate commitment.

Results indicated as “part 1” are included in De Simone et al., J Immunol, 2019 (DOI: https://doi.org/10.4049/jimmunol.1901072). Originally published in *The Journal of Immunology.* De Simone, G. *et al.* CXCR3 Identifies Human Naive CD8 + T Cells with Enhanced Effector Differentiation Potential . *J. Immunol.* **203**, 3179–3189 (2019). Copyright © [2019] The American Association of Immunologists, Inc.[104]

Results indicated as “part 2” are included in a Galletti*, De Simone*, Mazza*, et al., Nat Immunol, 2020 (DOI: https://doi.org/10.1038/s41590-020-0791-5 (*shared-first authors). Material from: Galletti, G. *et al.* Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans. *Nat. Immunol.* **21**, 1552–1562 (2020).[105]

3 Materials and methods

3.1 Cells

Peripheral blood mononuclear cells (PBMCs) were isolated from buffy coats from healthy donors by Ficoll gradient separation and used fresh for most of the assays. In some cases, PBMCs, previously frozen in liquid nitrogen in a suspension containing FBS+10% DMSO, were used. Surgically-removed lymph nodes were isolated as previously described[106]. Total $CD8^+$ T cells or naïve $CD8^+$ T cells were enriched by magnetic separation by using EasySep™ Human $CD8^+$ T Cell Isolation Kit (Stem Cell Technologies) or MojoSort Human $CD8^+$ Naive T Cell Isolation Kit (BioLegend), respectively, according to the manufacturers' instructions.

Cohort of samples (1st part)		
Parameter	**Subdivision**	
Sex (no.)	female	12
	male	42
	unknown	4
Age (yr.)	mean ± SD	40 ± 13
	range min - max	23 - 73
	unknown (no.)	10
Condition (no.)	healthy (PB)	54
	head and neck cancer patients**	4
Cohort of samples (2nd part)		
Parameter	**Subdivision**	
Sex (no.)	female	24
	male	95
	unknown	1
Age (yr.)	mean ± SD	50 ± 11
	range min - max	22 - 77
	unknown (no.)	3
Condition (no.)	healthy (PB)	100
	HIV+ (PB)	2
	non-small cell lung cancer patients*	6
	head and neck cancer patients**	6
	healthy bone marrow donors***	6

Table 1: Donors used in the study. Abbreviations: no.= number, yr.= years, SD= standard deviation. *donating PB and a tumor-free lung sample, **donating PB and a tumor-free lymph node, ***donating PB and bone marrow

3.2 Flow cytometry and cell sorting

Fluorochrome-conjugated monoclonal antibodies were purchased from BD Biosciences, BioLegend and eBioscience, and titrated to determine optimal concentrations[107]. Chemokine receptors expression was measured by incubating cells at 37°C for 20 min. Surface markers were measured by incubating cells at RT for 20 min. The Cytofix/Cytoperm kit (BD Biosciences) was used to detect intracellular cytokine expression. In all assays, cells were stained for 15 min at room temperature with Zombie Aqua fixable viability die to remove dead cells (BioLegend). T_NR3^- and T_NR3^+ were FACS-sorted according to the gating strategy depicted in Fig 2A. T_{PEX} and T_{STEM} were FACS-sorted according to the gating strategy depicted in Fig19C.

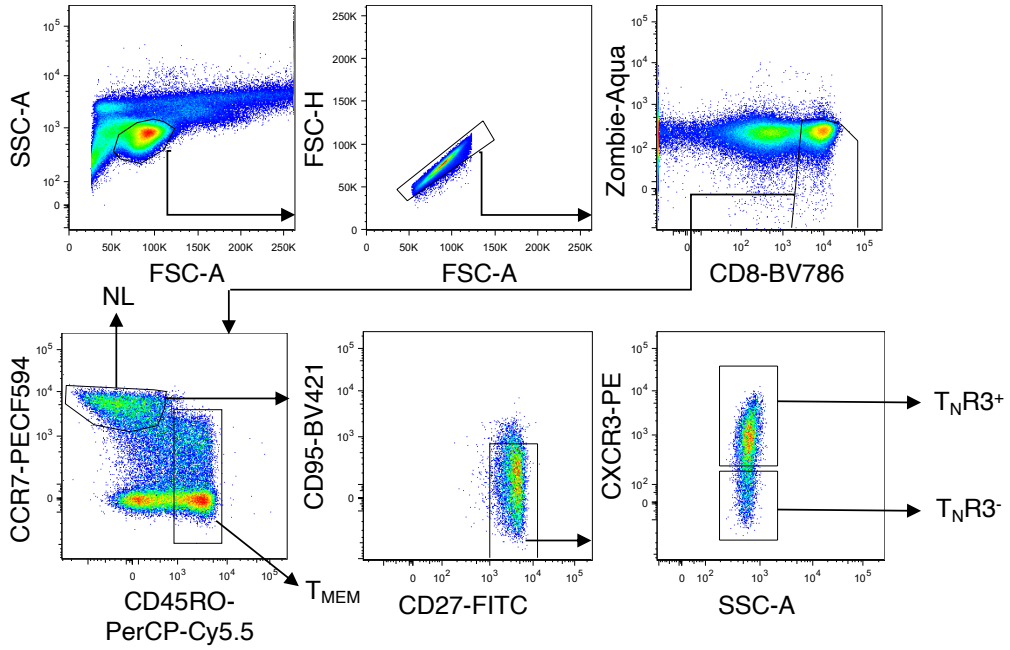


Figure 2: Gating strategy for the identification and FACS sorting of the cell subsets in the first part

All samples were acquired on LSR Fortessa or FACSSymphony A5 flow cytometers or separated via a FACSAria III cell sorter (all from BD Biosciences). Flow cytometry data were analyzed and

compensated with FlowJo 9 (FlowJo LLC) by using single-stained controls prepared with antibody-capture beads (BD Biosciences), as described[88].

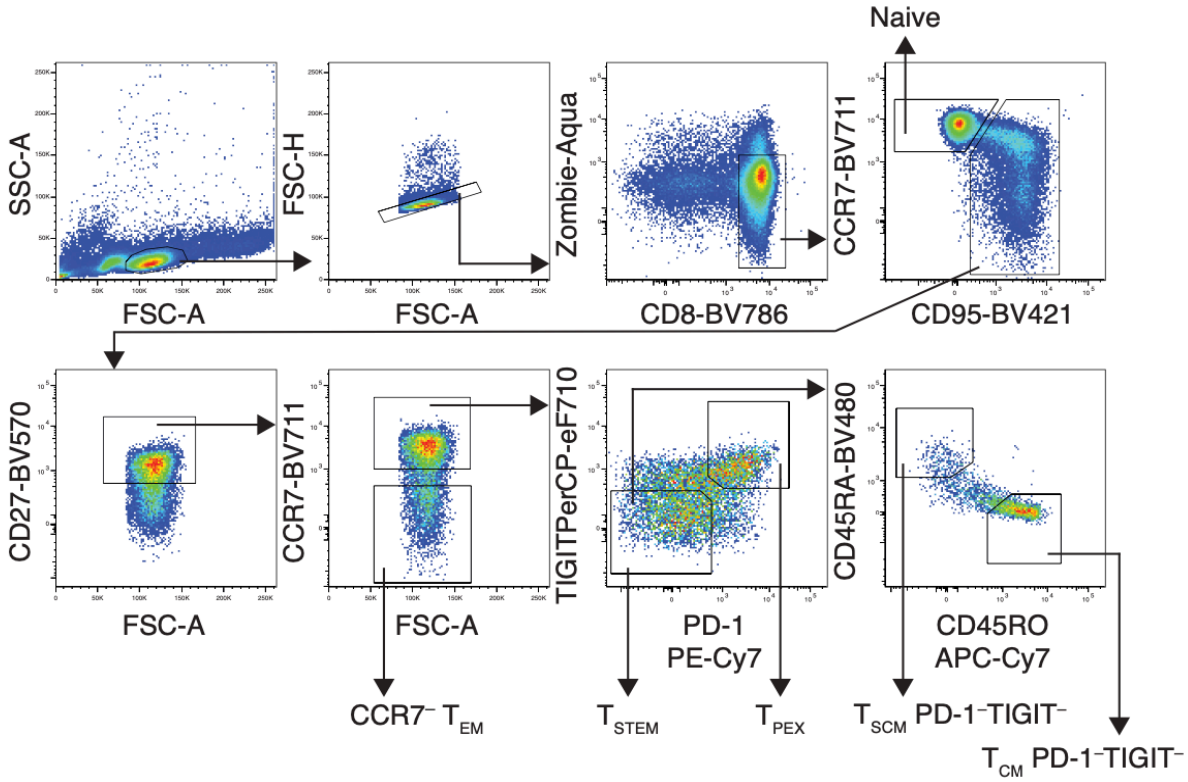


Figure 3: Gating strategy for the identification and FACS sorting of the cell subsets in the second part

ANTIBODY	FLUOROCHROME	CLONE	SOURCE	IDENTIFIER	RRID
Anti-human CD3	PE-Cy5	UCHT2	BD Biosciences	Cat # 555341	RRID:AB_395747
Anti-human CD5	BV605	UCHT2	BD Biosciences	Cat-# 563945	RRID:AB_2738500
Anti-human CD27	FITC	M-T271	BD Biosciences	Cat # 555440	RRID:AB_395833
Anti-human CD27	APC efluor780	O323	eBioscience	Cat # 47-0279-42	RRID:AB_1272040
Anti-human CD45RA	BV421	HI-100	BD Biosciences	Cat # 562885	RRID:AB_2737864
Anti-human CD45RO	PerCP-Cy5.5	UCHL1	Biolegend	Cat # 304222	RRID:AB_2174124
Anti-human CD73	BV605	AD2	BD Biosciences	Cat # 563199	RRID:AB_2738063
Anti-human CD95	PE-Cy5	DX2	BD Biosciences	Cat # 559773	RRID:AB_397317
Anti-human CD4	BV570	RPA-T4	Biolegend	Cat # 300534	RRID:AB_2563791
Anti-human CXCR3	PE	G025H7	Biolegend	Cat # 353706	RRID: AB_10962912
Anti-human CXCR3	BV711	G025H7	Biolegend	Cat # 353732	RRID:AB_2563533
Anti-human CXCR3	BV711	1C6	BD Biosciences	Cat # 563156	RRID:AB_2738034
Anti-human CXCR3	AF647	G025H7	Biolegend	Cat # 353712	RRID:AB_10962948
Anti-human CD95	BV605	DX2	Biolegend	Cat # 305627	RRID:AB_2562444
Anti-human CD45RA ()	FITC	ALB11	Beckman Coulter	Cat # A07786	N/A
Anti-human CD4 ()	PE-Cy5	13B8.2	Beckman Coulter	Cat # A07752	N/A
Anti-human CD19	PE-Cy5	J3-119	Beckman Coulter	Cat # A07771	N/A
Anti-human CD56	PE-Cy5	N901	Beckman Coulter	Cat # A07789	N/A
Anti-human TNF-α	PE	MAb11	eBioscience	Cat # 12-7349-81	RRID:AB_466207
Anti-human IFN-γ	APC-Cy7	4S.B3	Biolegend	Cat # 502530	RRID:AB_10663412
Anti-human Perforin	APC	B-D48	Biolegend	Cat # 353312	RRID:AB_2571969
Anti-human Granzyme B	Pacific Blue	GB11	Biolegend	Cat # 515408	RRID:AB_2562196
Anti-human MIP-1α	PE	CR3M	eBioscience	Cat # 12-9706-41	RRID:AB_2572721
Anti-human MIP-1β	AF700	D21-1351	BD Biosciences	Cat # 561278	RRID:AB_10612008
Anti-human CD127	PE	hIL-7R-M21	BD Biosciences	Cat # 557938	RRID:AB_2296056
Anti-human CCR7	BV421	G043H7	Biolegend	Cat # 353208	RRID:AB_11203894
Anti-human CD45	PE-Cy7	HI30	BioLegend	Cat # 304016	RRID:AB_314404
Anti-human CD8	BUV 805	SK1	BD Biosciences	Cat # 564912	RRID:AB_2744465
Anti-human CD8	BV786	RPA-T8	BD Biosciences	Cat # 563823	RRID:AB_2687487
Anti-human CD4	FITC	M-T477	BD Biosciences	Cat # 556615	RRID:AB_396487
Anti-human CCR7	BV711	G043H7	BioLegend	Cat # 353228	RRID:AB_2563865
Anti-human CD45RA	BV480	HI100	BD Biosciences	Cat # 566114	RRID:AB_2739516
Anti-human CD45RO	APC-H7	UCHL1	BD Biosciences	Cat # 561137	RRID:AB_10562194
Anti-human CCR7	BB660	150503	BD Biosciences	Cat # 625454	N/A
Anti-human CD95	BV421	DX2	BioLegend	Cat # 305624	RRID:AB_2561830
Anti-human CD95	BUV563	DX2	BD Biosciences	Cat # 624284	N/A
Anti-human CD95	APC	DX2	BD Biosciences	Cat # 558814	RRID:AB_398659
Anti-human CD45RO	BUV395	UCHL1	BD Biosciences	Cat # 562491	N/A
Anti-human CD3	BV650	OKT3	BioLegend	Cat # 317324	RRID:AB_2563352
Anti-human CD4	BV570	RPA-T4	BioLegend	Cat # 300534	RRID:AB_2563791
Anti-human CD4	BUV615	SK3	BD Biosciences	Cat # 624297	N/A
Anti-human CD3	BUV496	UCHT1	BD Biosciences	Cat # 564809	RRID:AB_2744388
Anti-human CD27	BV570	O323	BioLegend	Cat # 302825	RRID:AB_11149686
Anti-human TNF-α	APC-Cy7	mAB11	BioLegend	Cat # 502944	RRID:AB_2562870
Anti-human IFN-γ	PE-Cy7	B27	BioLegend	Cat # 506518	RRID:AB_2123321
Anti-human IL-2	APC	MQ1-17H12	BD Biosciences	Cat # 554567	RRID:AB_398571
Anti-human CD45	Pacific Blue	HI30	BioLegend	Cat # 304022	RRID:AB_493655
Anti-human CD3	PE-Cy5	HIT-3A	BD Biosciences	Cat # 555341	RRID:AB_395747
Anti-human CCR7	PE-CF594	150503	BD Biosciences	Cat # 562381	RRID:AB_11153301
Anti-human CD27	PE	M-T271	BD Biosciences	Cat # 560985	RRID:AB_10563213
Anti-human CD28	BV785	CD28.2	BioLegend	Cat # 302950	RRID:AB_2632607
Anti-human CD25	APC-R700	2A3	BD Biosciences	Cat # 565106	RRID:AB_2744339
Anti-human CD69	BUV737	FN50	BD Biosciences	Cat # 564439	RRID:AB_2722502
Anti-human CD107a	PE-Cy5	H4A3	BD Biosciences	Cat # 555802	RRID:AB_396136
Anti-human CD107a	BB630	H4A3	BD Biosciences	Cat # 624294	N/A
Anti-human CD127	PE-Cy5	eBioRDR5	eBioscience	Cat # 15-1278-42	RRID:AB_2043801
Anti-human PD-1	BV480	EH12.1	BD Biosciences	Cat # 566112	RRID:AB_2739514
Anti-human PD-1	PE-Cy7	EH12.2H7	BioLegend	Cat # 329918	RRID:AB_2159324
Anti-human TIGIT	PerCP-eFluor 710	MBSA43	eBioscience	Cat # 46-9500-42	RRID:AB_10853679
Anti-human TIGIT	FITC	MBSA43	eBioscience	Cat # 11-9500-41	RRID:AB_2572529
Anti-human TIGIT	BV421	A15153G	BioLegend	Cat # 372710	RRID:AB_2632925
Anti-human HLA-DR	BUV661	G46-6	BD Biosciences	Cat # 565073	RRID:AB_2722500
Anti-human CD38	BV711	HIT2	BioLegend	Cat # 303528	RRID:AB_2563811
Anti-human CD103	BV421	Ber-ACT8	BioLegend	Cat # 350213	RRID:AB_2563513
Anti-human CD161	BV605	HP-3G10	BioLegend	Cat # 339916	RRID:AB_2563607
Anti-human CD14	BV510	M5E2	BioLegend	Cat # 301842	RRID:AB_2561946
Anti-human Granulysin	Alexa Fluor 488	RB1	BD Biosciences	Cat # 558254	N/A
Anti-human EOMES	PE-eFluor 610	WD1928	eBioscience	Cat # 61-4877-41	RRID:AB_2574615
Anti-human Granzyme B	APC-R700	GB11	BD Biosciences	Cat # 561016	RRID:AB_2033973
Anti-human Granzyme K	PE	GM6C3	Santa Cruz	Cat # sc-56125 PE	RRID:AB_2263772
Anti-human Granzyme K	Alexa Fluor 647	GM6C3	Santa Cruz	Cat # sc-56125 AF6	RRID:AB_2263772
Anti-human IRF4	Alexa Fluor 488	IRF4.3E4	BioLegend	Cat # 646406	RRID:AB_256326
Anti-human IRF8	APC	V3GYWCH	eBioscience	Cat # 17-9852-80	RRID:AB_2573317
Anti-human T-bet	PE-Cy7	4-B10	eBioscience	Cat # 25-5825-82	RRID:AB_11042699
Anti-human LEF1	PE	C12A5	Cell Signaling	Cat # 14440	N/A
Anti-mouse CD45.1	PE-Cy7	A20	BD Biosciences	Cat # 560578	RRID:AB_1727488
Anti-mouse CD45	PerCP-Cy5.5	30-F11	BioLegend	Cat # 103132	RRID:AB_893340
Anti-mouse CD45.1	BV605	A20	Biolegend	Cat # 110738	RRID:AB_2562565
Anti-mouse CD45.2	APC-Cy7	104	Biolegend	Cat # 109824	RRID:AB_830789
Anti-mouse CD8	BV570	53-6.7	Biolegend	Cat # 100740	RRID:AB_2563055
Anti-mouse CD62L	APC	MEL-14	BD Biosciences	Cat # 553152	RRID:AB_398533
Anti-mouse CD44	FITC	IM7	eBioscience	Cat # 11-0441-82	RRID:AB_465045
Anti-mouse CD3	BV650	145-2C11	BD Biosciences	Cat # 564378	RRID:AB_2738779
Anti-mouse CD127	BV786	SB/199	BD Biosciences	Cat # 563748	RRID:AB_2738403
Anti-mouse KLRG1	BV421	2F1/KLRG1	Biolegend	Cat # 138414	RRID:AB_2565613
Anti-mouse CXCR3	PE	CXCR3-173	Biolegend	Cat # 126506	RRID:AB_1027650

Table 2: List of Flow Cytometry antibodies used in the study

3.3 Age-associated changes of T_N cell subsets

T_NR3^- and T_NR3^+ cells were quantified in a cohort of 1,938 individuals from the general population (815 males and 1123 females), aged 19-105, belonging to the SardiNIA Study[108,109]. Immunophenotyping was carried out by flow cytometry on fresh blood samples. To avoid circadian fluctuations and time dependent artefacts, blood samples were collected in heparin tubes at 8 am and immune-phenotypes were performed in the same recruitment center within two hours from the withdrawal. Blood was then antibody-stained, erythrocyte lysed and acquired with a FACSCAriaIII analyzer (BD Biosciences). $CXCR3^+$ T cells were quantified in the naïve-like CD8 population, defined as $CD3^+CD4^-$ $CD45RA^+CCR7^+CD127^+CD161^-PD\text{-}1^-$.

3.4 Cell culture

Cells were cultured in RPMI 1640 medium supplemented with 10% FBS, 1% penicillin/streptomycin and 2mM L-glutamine (hereafter referred to as R10). To induce cytokine production, FACS-sorted T_NR3^- and T_NR3^+ and total PBMCs were stimulated in a final volume of 200 μL with phorbol 12-myrystate 13-acetate (PMA; 10 ng/mL) and ionomycin (500 ng/mL) (both from SigmaAldrich) for 6 hours in the presence of the protein transport inhibitors GolgiPlug (1μL/mL of cell culture, BD Biosciences) and GolgiStop (0.67μL/mL of cell culture, BD Biosciences), according to manufacturer's instructions.

Cells were also stimulated or not with Staphylococcus Enterotoxin-B (SEB; 1μg/mL; SigmaAldrich) overnight. To evaluate differentiation and proliferation capability, cells were stimulated with anti-CD3/CD28

activation beads (Dynabeads; Thermo Fisher Scientific) in combination with cocktail of human cytokines, according to experimental design. Human cytokines for cell cultures (Peprotech) were pre-titrated and then used at 10 ng/ml (TGF-β, IL-2, IL-12, IL-7, IL-15) for 3 to 4 days. To evaluate memory maintenance in response to homeostatic cytokines, cells were stimulated with IL-15 alone (25 ng/ml) for 10 days.

3.5 Quantification of TRECs

T_NR3^-, T_NR3^+, T_{SCM} and bulk $CD45RO^+$ memory T cells were FACS-sorted in PBS without Ca^{2+} and Mg^{2+} ($PBS^{-/-}$) then washed twice in $PBS^{-/-}$ and frozen at -80°C. After thawing, cells were lysed in a Proteinase K solution (Roche; 100 μg/mL diluted in 10mM Tris-HCl pH8; 10μL/100000 cells). TREC content was determined by quantitative real-time PCR, as previously described[110], and normalized to the number of cells contained in each sample, determined via quantification of *FAS*.

3.6 Quantification of gene expression via qPCR

Total RNA, purified with RNeasy Micro Kit (Qiagen) with DNAse (Qiagen), was retro-transcribed using High capacity cDNA Reverse Transcription Kit (Applied Biosystems) and analyzed by qPCR with hydrolysis probes *CXCR3* (Hs00171041_m1). Quantification of *B2M* (Hs00187842_m1) served as reference gene. qPCR was performed with Universal PCR Master Mix, No Amperase UNG (Roche) in MicroAmp Fast Optical 96-Well reaction Plate (Applied Biosystems) on the ABI 7900HT Sequence Detection System (Applied Biosystems). Expression levels were

normalized (∆Ct) to *B2M* endogenous control according to the formula $2^{-(Ct\ CXCR3 - Ct\ B2M)}$.

3.7 Enumeration of antigen-specific CD8$^+$ T cells by MHC class I tetramers

Biotinylated HLA-A*0201/CMV pp65$_{495-503}$ NLVPMVATV (NV9), HLA-A*0201/Flu matrix protein$_{58-66}$ GILGFVFTL (GL9), HLA-A*0201/MART-1$_{26-35}$ ELAGIGILTV, and HLA-A*0201/CMV KA$_{227/8}$ NLVPMVATV (KA) monomers were multimerized with streptavidin-BV421 (Sigma), PE (Sigma) and APC (Life Technologies), as previously described[111]. Cells were stained for 15 min at 37°C (1 µg tetramer in 100 µl). On average, 6x10^6 PBMCs were acquired by flow cytometry.

3.8 Enumeration of antigen-specific T_N cell precursors by libraries of amplified T cells

Peripheral Blood Mononuclear Cells (PBMCs) were isolated with Ficoll-Paque Plus (GE Healthcare). Monocytes and total CD8$^+$ T cells were isolated by positive selection using CD14 and CD8 magnetic microbeads, respectively (Miltenyi Biotech). Two subsets of CD8$^+$ CD45RA$^+$ CCR7$^+$ CD27$^+$ CD95$^-$ CD4$^-$ CD19$^-$ CD56$^-$ naïve T cells were sorted from total CD8$^+$ T cells with a FACS Aria, on the basis of CXCR3 expression. Memory CD8$^+$ T cells were sorted as control subset.

T cells were cultured in RPMI 1640 medium supplemented with 2 mM glutamine, 1% (vol/vol) nonessential amino acids, 1% (vol/vol) sodium pyruvate, penicillin (50 U/ml), streptomycin (50 µg/ml) (all from Invitrogen) and 5% human serum (Swiss Red Cross). The sorted T cells (2,000

cells/well) were polyclonally stimulated with 1 µg/ml PHA (Remel) in the presence of irradiated (45Gy) allogeneic feeder cells (2.5×10^4 per well) and IL-2 (500 IU/ml), in a 96-well plate. T cell lines were expanded as previously described[112]. Library screening was performed 14-21 days after initial stimulation, by culturing thoroughly washed T cells (2.5×10^5/well) with autologous irradiated B cells (2.5×10^4), with or without a three-hour pulse with different antigens. The antigens used included the HIV-1 peptide pool (1 µg/ml/peptide, comprising 386 18-mer peptides spanning the entire 2004 consensus clade C HIV-1 proteome), the Zika virus H/PF/2013 peptide pool (1 µg/ml/peptide, comprising 669 10-mer peptides spanning Env, NS3 and NS5 proteins), the human cytomegalovirus (CMV) peptide pool (1 µg/ml/peptide, from A. Sette's laboratory, La Jolla Institute for Immunology, comprising 198 peptides of length 8-11aa), the Epstein–Barr virus (EBV) peptide pool (1 µg/ml/peptide, from A. Sette's laboratory, LJI, comprising 218 peptides of length 8-11aa), the Influenza A virus A/California/07/2009 (H1N1) peptide pool (2 µg/ml/peptide, comprising 351 15-mer peptides spanning Hemagglutinin, Neuraminidase, Matrix protein 1 and Nucleoprotein). Proliferation was assessed on day 4, after incubation for 16 h with 1 µCi/ml [methyl-^{3}H]thymidine (Perkin Elmer). Precursor frequencies were calculated based on numbers of negative wells, assuming a Poisson distribution, and are expressed per million cells.

3.9 Antigen-specific T cell proliferation and effector functions

CMV seronegative buffy coat donors were identified by coating plates with 2 µg/ml sonicated HCMV, followed by ELISA detection of plasma IgGs. Sorted T_NR3^-, T_NR3^+ and T_{MEM} $CD8^+$ T cells from these donors were labeled with CFSE and cultured at a ratio of 2:1 with irradiated autologous monocytes pre-pulsed for 5h with human CMV lysate or seasonal influenza virus vaccine (Influvac 2017/2018, from Mylan) together with CMV peptide pool or Influenza A Matrix protein 1 peptide pool, respectively. On day 10, T cells were stimulated with PMA and ionomycin for 5h in the presence of brefeldin A for the last 2h (all reagents from Sigma-Aldrich). Cell viability was determined by staining with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (ThermoFisher), according to the manufacturer's instructions. Subsequently cells were fixed and permeabilized with Cytofix/Cytoperm (BD Biosciences), and then stained for flow cytometric analysis.

3.10 TCR deep sequencing and data analysis

T_NR3^-, T_NR3^+ and $CD45RO^+$ T_{MEM} were FACS-sorted in triplicate (300,000 cells/subset) in 1.2 mL RLT buffer (Qiagen; final dilution <20%). T_{PEX}, T_{STEM} and T_{EM} were FACS-sorted in duplicate (150,000 cells/subset). RNA was extracted using RNeasy mini kit (Qiagen) according to the manufacturer protocol. Unique molecular identifier (UMI)-labelled 5'RACE TRB cDNA libraries were prepared using Human TCR Profiling Kit (MiLaboratory LLC). All extracted RNA was used for cDNA synthesis, and all synthesized cDNA was used for PCR amplification. Libraries were

prepared in parallel using the same number of PCR cycles and sequenced 150+150 bp on Illumina NextSeq. We obtained about 135 mln TRB sequencing reads (1.5 mln ± 0.3 mln reads per library), from which about 4 mln unique UMI-labelled TRB cDNA molecules (53,000 ± 10,000 molecules per library) were extracted using MIGEC[113] and MiXCR[114] software with MIGEC threshold set as at least 2 sequencing reads per UMI. Each library contained an average of 40,000±10,000 functional (in-frame, without stop-codons) CDR3 nucleotide sequence variants (unique TRBβ clonotypes). Averaged physico-chemical characteristics of 5 amino acid residues located in the middle of TRBβ CDR3 (weighted by clonotype size) were analyzed using VDJtools software[115]. In particular, we analyzed averaged statistical potential of CDR3:epitope interactions (the estimated "energy" of interaction between CDR3 and peptide[116], "strength" of interaction (derivative of "energy"), hydrophobicity (Kidera factor 4)[117,118], and "volume" indexes. Diversity metrics were analyzed by VDJtools after downsampling to the same number of 5,000 (first project part) and 3,000 (second project part) randomly selected UMI-labeled TRB cDNA molecules per sample.

3.11 Gene expression profiling by DNA microarrays and data analysis

Sorted T_NR3^-, T_NR3^+ and $CD45RO^+$ T_{MEM} cells were washed twice in $PBS^{-/-}$, resuspended in RLT buffer (Qiagen), processed and analyzed as previously described[119]. Briefly, microarray probe fluorescence signals were converted to expression values using robust multiarray average (RMA) procedure[120] of Bioconductor Affy package. Specifically, fluorescence intensities were background adjusted and normalized using

quantile normalization. Log_2 expression values for a total of 32,500 custom probe sets were calculated using median polish summarization and custom chip definition files for Affymetrix Human Transcriptome Array 2.0 based on Entrez genes (hta20_Hs_ENTREZG version 21.0.0). All data analyses were performed in R version 3.4.4. Differentially expressed genes between human T_NR3^- and T_NR3^+ cells were identified by the limma algorithm coded in the same R package[121].

3.12 Pathway analysis of microarray data

Pathway analysis was performed using GSEA software (http://software.broadinstitute.org/gsea/msigdb/) and gene sets from Molecular Signatures Database (Version 6.2). Namely, we used gene sets of the c2 (c2.cp.reactome.v6.2) and immunological signatures (c7.all.v6.2) collections. GSEA was applied on log_2-transformed expression data of T_NR3^- and T_NR3^+ cells. Prior to this analysis, mouse gene IDs obtained from the comparison between T_NR3^+ vs. $CD5^{lo}$ or $CD5^{hi}$ cells were converted into the corresponding human orthologue genes using the HUGO Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/cgi-bin/hcop). Gene sets were considered significantly enriched at FDR≤0.05 when using Signal2Noise as metric and 1,000 permutations of the gene sets.

3.13 Mice

Animal protocols were approved by the Humanitas IACUC and by the Italian Ministry of Health (No. 452/2018-PR), and by the San Raffaele Scientific Institute Ethical Committee and by the Italian Ministry of Health (No. 282/2015-PR).

3.14 Mouse $CD8^+$ T_N cell sorting and RNAseq

Spleens from C57BL/6 male mice (Charles River Laboratories) were collected at 12 weeks of age and mechanically smashed by plunger of a 2 ml syringe with a 40 μm cell strainer. Splenocytes were enriched for $CD8^+$ T cells with the MojoSort mouse $CD8^+$ T cell isolation kit (Biolegend) and stained for FACS. FACS-sorted cells were lysed with 50 μL RLT buffer (Qiagen), with the addition of 1 μL RNAse inhibitor (Qiagen). RNA extraction was performed with the RNeasy Micro Kit (Qiagen), according to the manufacturer's instructions. RNAseq libraries were prepared by using the SMARTseqv4 RNA kit for sequencing (Clontech-Takara) which is designed to generate high-quality, full-length cDNA directly from 10 pg to 10 ng of total RNA. This Kit incorporates Clontech's proprietary SMART® (Switching Mechanism at 5' End of RNA Template) technology. This technology relies on the template switching activity of reverse transcriptases to enrich for full-length cDNAs and to add defined PCR adapters directly to both ends of the first-strand cDNA. This ensures the final cDNA libraries contain the 5' end of the mRNA and maintain a true representation of the original mRNA transcripts; these factors are critical for transcriptome sequencing and gene expression analysis. Libraries have been pooled at equimolar concentrations and sequenced on an Illumina NextSeq 500 platform; at least 20 million 75bp single end reads per sample have been generated.

3.15 RNAseq data analysis

After quality control with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), 75-bp single-end reads were aligned to the *Mus musculus* reference genome (Ensembl assembly GRCm38) using STAR (version 2.5.1b)[122]. Alignment was performed with default parameters, and reads associated to annotated genes were counted with HTSeq and "*-quantmode TranscriptomeSAM GeneCounts"* options. Differential expression between murine T_N subsets was assessed using the edgeR package (version 3.22)[123] applying Benjamini-Hochberg multiple testing correction to estimate the FDR.

3.16 In vivo studies

LCMV-P14 TCR transgenic mice were obtained through the Swiss Immunological Mutant Mouse Repository (Zurich, Switzerland). $CD8^+$ T_N cell subsets from CD45.1 LCMV P14 mice (kept in SPF conditions) were isolated as described above. 1-2x10^4 cells were injected i.v. into CD45.2 WT recipient animals (C57BL/6) 16 hours prior i.v. infection with 2x10^5 FFU of LCMV Armstrong. Animals were sacrificed at day 7 and 14 after infection, single cell suspensions were obtained from spleens and stained for flow cytometry as described above.

For in vivo persistence experiments, Eight-week-old female JAX NOD.Cg-$Prkdc^{scid}$ $Il2rg^{tm1Wjl}$/SzJ (NSG, #005557, Charles River) mice were infused by retroorbital injection with T_{STEM}, T_{MEX} or T_{EM} cells (1x10^6 per mouse), freshly sorted from healthy donors' PBMCs (n=2). A third experiment was excluded from the analysis due to poor recovery of cells following transfer

in primary recipients. Autologous $CD8^-$ PBMCs ($6x10^6$ per mouse) were co-injected as support. Blood sampling from the lateral tail vein was performed at 11, 18 and 28 days post transfer to monitor T cell engraftment and expansion. Mice were sacrificed at day 28. Spleens were collected from mice, processed to single cell suspension and the number of human T cells was determined. Absolute numbers of T cells in the blood was determined by CountBright Absolute Counting Beads (Invitrogen), according to manufacturer's protocol. The frequency of human $CD4^+$ and $CD8^+$ T cell subsets was determined by flow cytometry. To maximize the recovery of human T cells, spleens and lungs from the same experimental group were mixed, the CD4:CD8 ratio of T_{STEM} and T_{MEX} normalized with autologous $CD8^-$ PBMCs and injected as above in secondary NSG recipients ($1x10^6$ $CD8^+$ T cell per mouse). T_{EM} cells could not be transferred due to low cell recovery. Expansion and absolute numbers of T cells were monitored as mentioned above.

For tumor experiments, cell were transduced as follows: T cell subsets were FACS-sorted from total PBMCs of one healthy donor. $CD3^+$ T cells, used as a control, were purified by magnetic separation. T cells were stimulated with the MACS-GMP T Cell TransAct cocktail (Miltenyi), transduced on day 1 with a bidirectional lentiviral vector encoding a CD19 CAR harboring CD28 costimulus in sense and the LNGFR marker gene in antisense, and finally kept in culture for 13 days in TexMacs medium (Miltenyi) supplemented with 3% FBS, 1% penicillin/streptomycin and IL-7/IL-15 (Miltenyi). Eight-week-old female NSG mice were intravenously injected with $0.5x10^6$ $Lucia^+/NGFR^+$/NALM-6 cells, followed 4 days later by injection of $3x10^6$ CAR19-redirected T_{STEM}, T_{MEX} or total $CD3^+$ T cells. Untransduced $CD3^+$ T cells were used as controls. Tumor progression

was monitored weekly by bioluminescence detection by using the QUANTI-Luc detection reagent (InvivoGen) and expressed as relative light units (RLUs), according to the manufacturer's instructions.

3.17 tSNE analysis of high-dimensional CyTOF data

Public CyTOF data of T cell antigen expression from tonsils, spleen, liver, gut, skin and lung were previously described in Wong et al.[124] and were downloaded from https://flowrepository.org/. Files (debarcoded samples) were imported in FlowJo and concatenated into a single .fcs file (~2,500 events/sample) which was subsequently subjected to t-SNE analysis (Barnes-Hut implementation) with the following parameters: iterations, 1000; perplexity, 40; initialization, deterministic; theta, 0.5; eta: 200. All markers listed in Fig 3A , except CXCR3, were included in the analysis.

3.18 Single-cell RNA sequencing

100,000 FACS-sorted $CD95^{+}CD8^{+}$ T cells from four healthy donors were resuspended in 1 ml $PBS^{-/-}$ plus 0.04% BSA and washed two times by centrifugation at 450 rcf for 7 min. After the second wash, cells were resuspended in 100 µL and counted with an automatic cell counter (Countess II, Thermo Fisher) to get a precise estimation of total number of cells recovered. Afterwards, about 20,000 cells of each sample were loaded into one channel of the Single Cell Chip A using the Single Cell 3' v2 reagent kit (10X Genomics) for Gel bead Emulsion generation into the Chromium system. Following capture and lysis, cDNA was synthesized and amplified for 14 cycles by following the manufacturer's protocol (10X Genomics). 50 ng of the amplified cDNA for each sample were then used

to construct sequencing libraries. Sequencing was performed on the NovaSeq 6000 Illumina system following 10X Genomics instruction for reads generation. An average sequencing depth of at least ~50,000 reads/cell was obtained for each sample.

3.19 Single-cell RNA sequencing analysis

Sample demultiplexing, barcode processing and UMI counting were performed by using the 10x Genomics pipeline Cell Ranger v2.1.1 (https://support.10xgenomics.com). Briefly, raw base call files were demultiplexed in FASTQ format using the “cellranger mkfastq” pipeline. Then, cellranger count was run with --transcriptome=refdata-cellranger-GRCh38-1.2.0 for each sample. The outputs of cellranger count for individual samples were concatenated into one matrix and the libraries were normalized to the same sequencing depth using the “cellranger aggr” pipeline. The concatenated gene-cell barcode matrix was loaded into R by using the Seurat package[125]. To exclude genes that might be detected from random noise, we filtered genes whose expression was detected in less than three cells, and retained cells where at least 200 features were detected. Moreover, we considered outliers and filtered out cells if they had unique feature counts over 3,500 or less than 200 and more than 10% of mitochondrial counts. The resulting dataset was normalized through a global-scaling method, converted by a scale factor (10,000 by default) and log-transformed using the “ScaleData” Seurat implemented function. The resulting data were subjected to clustering analysis using standard Seurat package procedures. In particular, the first 20 principal components and a resolution equal to 0.6, were used to cluster the cells into subpopulations through a graph-based unsupervised clustering approach implemented in

Seurat (“FindClusters” function). Differentially expressed genes of each single cluster compared to all other cells were identified using the “FindAllMarkers” function. All markers distinguishing cluster 2 from cluster 6, were identified by using the “FindMarkers” function coded in Seurat package.

3.20 High-dimensional flow cytometry data analysis

Flow Cytometry Standard (FCS) 3.0 files were imported into FlowJo software version 9, analyzed by standard gating to remove aggregates and dead cells, and CD95+ bulk memory CD8+ T cells were identified. 5,000 CD95+ T cells per sample were subsequently imported in FlowJo version 10, biexponentially transformed and exported for further analysis in Python (version 3.7.3) by a custom-made script that makes use of PhenoGraph (originally retrieved from the scikit-learn package; full script available at https://github.com/luglilab/Cytophenograph). Lymph nodes, bone marrow and adjacent cancer-free lung tissues were labelled with a unique computational barcode for further identification and converted in comma separated (CSV) files and concatenated in a single matrix by using the merge function of pandas package. K value, indicating the number of nearest neighbors identified in the first iteration of the algorithm, was set equal to 1000 for clustering. UMAP was obtained by UMAP Python package and visualized in FlowJo 10.

3.21 Cytometry by Time of Flight (CyTOF) antibody and streptavidin labeling

Antibodies were labelled by following MAXPAR conjugation protocol available from Fluidigm. Antibodies were titrated for optimal concentrations before staining of donor samples. Streptavidin was expressed, refolded, and purified in-house, then labelled as described[126].

3.22 Preparation of peptide-MHC monomers

Myc-tagged HLA-A*02:01 was synthesized, refolded with appropriate UV-cleavable peptide, and biotinylated as described[127]. Biotinylated HLA-A*0201 with a UV-cleavable peptide was diluted to 100 µg/mL in PBS and mixed with 5 µL of 1 mM peptide for a total of 100 µL per reaction. Reactions were exposed to long-wave UV light (368 nm) for 5 mins in a Fisher UV Crosslinker, then rotated 180 degrees and exposed again. Plate was incubated overnight at 4^{o} C before tetramerization the next day.

3.23 Mass cytometry staining with combinatorial peptide-MHC tetramers and magnetic bead enrichment

Heavy-metal labelled streptavidin were mixed together according to a triple coding scheme as previously described (20 µL of 50 µg/mL dilution for a total of 60 µL per combination) (Ref 51, 52 paper). Combined streptavidin were then added to the peptide-MHC monomers, prepared as described above, in 3 separate additions of 20 µL with 5 minutes of incubation time in between each addition. Tetramerized peptide-MHC were then combined in a 50 kDa Amicon concentrator (Millipore) and concentrated at 4000xG, then diluted to 200 µL per donor and filtered

through a 0.1 micron Amicon centrifugal filter (Millipore). Cryopreserved PBMCs for three healthy donors and two HIV seropositive donors were obtained from HVTN in advance and thawed at 37oC before transferring into RPMI, then FACS buffer (0.5% BSA, 0.02% sodium azide in PBS). Cells were stained for one hour with tetramer cocktail at room temperature, along with CCR5, CCR7, TIGIT, PD-1, and CXCR5 to improve staining quality. Tetramer cocktail was washed off with MACS buffer (0.5% BSA, 2 mM EDTA), and ~2 million cells were separated as pre-enriched fractions. The remaining cells were incubated with a 1:3 dilution of anti-c-myc magnetic beads (Miltenyi Biotec) on ice for twenty minutes, then magnetically enriched. LS magnetic columns (Miltenyi Biotec) were washed with 2 mL MACS buffer, followed by application of cells. Flowthrough was collected with 3x2mL washes before magnetic columns were removed from the magnetic field and enriched cells were collected by washing the column with 5 mL of MACS buffer and a plunger. Cell numbers were counted and, when applicable, enriched fractions were supplemented with flowthrough cells to limit cell loss due to staining. Antibody staining, live/dead discrimination and DNA staining were performed as described (55) before acquisition on the CyTOF.

3.24 STELA assay and data analysis

T cells (6,000 cells/subset) were FACS-sorted in PBS–/–, washed twice in PBS–/– and frozen at -80°C. Then, DNA was extracted using a QIAamp DNA Micro Kit (Qiagen). Single telomere length analysis was carried out at the XpYp telomere as described previously[128]. Briefly, genomic DNA was eluted in 35 µL Tris (10 mM) to which 0.75 µL Telorette-2 linker (10 µM) was added. Multiple PCRs were then performed for each test DNA. Each

reaction was set up in a final volume of 10 µL containing 1 µL of DNA and 0.5 µM of the telomere-adjacent and Teltail primers in 75 mM Tris-HCl pH 8.8, 20 mM (NH4)2SO4, 0.01% Tween-20, and 1.5 mM MgCl2, with 0.5 U of a 10:1 mixture of Taq (Thermo Fisher Scientific) and Pwo polymerase (Roche Molecular Biochemicals). The reactions were processed in a Tetrad2 Thermal Cycler (Bio-Rad). DNA fragments were resolved via 0.5% Tris-acetate-EDTA agarose gel electrophoresis and identified via Southern hybridization with a random-primed anti-32P-labeled (PerkinElmer) TTAGGG repeat probe, together with probes specific for molecular weight markers at 1 kb (Stratagene) and 2.5 kb (Bio-Rad). Hybridized fragments were detected using a Typhoon FLA 9500 Phosphorimager (GE Healthcare). The molecular weights of the DNA fragments were calculated using a Phoretix 1D Quantifier (Nonlinear Dynamics).

3.25 Library preparation and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

Cells were FACS-sorted in 500 µL of R10. Freshly sorted 50,000 cells from each subset were washed in $PBS^{-/-}$ and resuspended in 50 µL of Lysis Buffer (10mM Tris-HCl pH 7.4, 10mM MgCl2, 0.1% Igepal CA-630). Nuclei were pelleted by centrifugation for 10 min at 500g, 4°C, and resuspended in 50 µL of reaction buffer containing 1 µl of Tn5 transposase (made in house), 10 µL of 5x transposase buffer (50mM Tris-HCl, pH 8.4 and 25mM MgCl2), and 39 µL of milliQ water. The reaction was incubated at 37°C for 30 min with mixing at 300 RPM, followed by the addition of 10 µL of clean-up buffer (900mM NaCl, 30mM EDTA), 5 µL of 20% SDS, 0.7 µL of milliQ water and 4.3 µL of Proteinase K (18.6 µg/µL) (Thermo Fisher

Scientific) and incubation for 30 min at 40°C. Tagmented DNA was isolated using SPRI beads (2x) and amplified by PCR. Fragments smaller than 600 bp were isolated by negative size selection (using 0.65x SPRI beads) and then purified with 1.8x SPRI beads. Libraries were qualitatively assessed by using TapeStation 4200 and quantified by Qubit Fluorimeter. Libraries were then multiplexed in an equimolar pool and sequenced on a NextSeq-500/550 Illumina Platform by generating between 20 and 50 million 75bp-SE reads per sample, depending on the experiment.

3.26 ATAC-sequencing analysis

Read quality was assessed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and low-quality bases and adaptors were trimmed by Cutadapt. Samples were aligned to GRCh38 reference genome with BWA mem with the default condition (v0.7.17)[129]. Mitochondrial reads were removed using SAMtools (v1.9) while PCR duplicated were filtered using MarkDuplicates function from Picard tool (v2.19) (http://broadinstitute.github.io/picard/). Open chromatin was detected with MACS2 (v2.1.2)[130] with a FDR lower than 0.01. The number of reads in each peak was determined with featureCounts, differentially accessible peaks were identified after DESeq2 (v1.20) normalization using a FDR cut-off lower than 0.05. Peaks were annotated with HOMER annotatePeaks.pl and scanned for de novo and known motifs using HOMER findMotifsGenome.pl (v4.9.1)[131].

3.27 Statistical analysis

Statistical analyses were performed using GraphPad Prism version 7 or R software version 3.4.3. P values were considered statistically significant when <0.05, unless specified otherwise. Statistical tests used in specific comparisons are indicated in Figure Legends

Chemicals, Peptides, and Recombinant Proteins		
Recombinant human IL-2	Peprotech	Cat # 200-02-A
Recombinant human IL-7	Peprotech	Cat # 200-07-10UG
Recombinant human IL-12	Peprotech	Cat # 200-12-10UG
Recombinant human IL-15	Peprotech	Cat # 200-15B
Phorbol 12-Myristate 13-Acetate	Sigma-Aldrich	Cat # P8139-1MG
Ionomycin Calcium Salt	Sigma-Aldrich	Cat # I0634
Foxp3/Transcription Factor Staining Buffer Set	eBioscience	Cat # 00-5523-00
CellTrace CFSE Cell Proliferation Kit	ThermoFisher	Cat # C34554
Zombie Aqua Fixable Viability kit	Biolegend	Cat # 423101
GolgiPlug Protein Transport Inhibitor	BD Biosciences	Cat # 555029
Proteinase K	Merck	Cat # 3115887001
Universal PCR Master Mix no Amperase UNG	ThermoFisher	Cat # 4324018
Streptavidin BV421	Biolegend	Cat # 405225
Streptavidin PE	Sigma	Cat # E4011-ML
Streptavidin APC	ThermoFisher	Cat # S32362
Ficoll-Paque Plus	GE Healthcare	Cat # 17-1440-03
Critical Commercial Assays		
MojoSort human CD8 naïve isolation kit	Biolegend	Cat # 480046
EasySep human CD8+ enrichment kit	Stem Cell Tech.	Cat # 19053
MojoSort mouse CD8 isolation kit	Biolegend	Cat # 480035
RNeasy Micro Kit	Qiagen	Cat # 74004
CD14 Microbeads, human	Miltenyi Biotech	Cat # 130-050-201
CD8 Microbeads, human	Miltenyi Biotech	Cat # 130-045-201
Experimental Models: Organisms/Strains		
C57BL/6	Charles River	
B6:D2-TCR LCMV (P14)	SWIMM	
Oligonucleotides		
TREC-F: CACATCCCTTTCAACCATGCT	Sigma	8021100303-000010
TREC-R: GCCAGCTGCAGGGTTTAGG	Sigma	8021100303-000020
TREC-PROBE : [6FAM]ACACCTCTGGTTTTTGTAAAGGTGCCCA(	Sigma	8021100303-000030
FAS-F: GGCTCTGTGAGGGATATAAAGACA	Sigma	8021100303-000040
FAS-R: CAAACCACCCGAGCAACTAATCT	Sigma	8021100303-000050
FAS-PROBE : [6FAM]CTGTTCCGTTTCCTGCCGGTGC[BHQ1]	Sigma	8021100303-000060
B2M:	ThermoFisher	Hs00187842_m1
CXCR3:	ThermoFisher	Hs00171041_m1
Recombinant DNA		
pGem T easy plasmid	Promega	N/A

Table 3: List of the reagents used in the study

4 Results

4.1 Results part 1

4.1.1 CXCR3 identifies subsets of T_N cells in humans

4.1.1.1 Phenotypical characterization of T_NR3^- and T_NR3^+

We aimed to analyze human T_N cells according to stringent phenotypic characterization, i.e., defined as $CD45RO^-$ $CCR7^+$ $CD27^+$ $CD95^-$ in order to exclude any possible memory contaminations. We found that CXCR3, a C-X-C motif chemokine receptor involved in migration towards inflamed tissues, was not uniformly expressed by $CD8^+$ T_N cells. We were able to

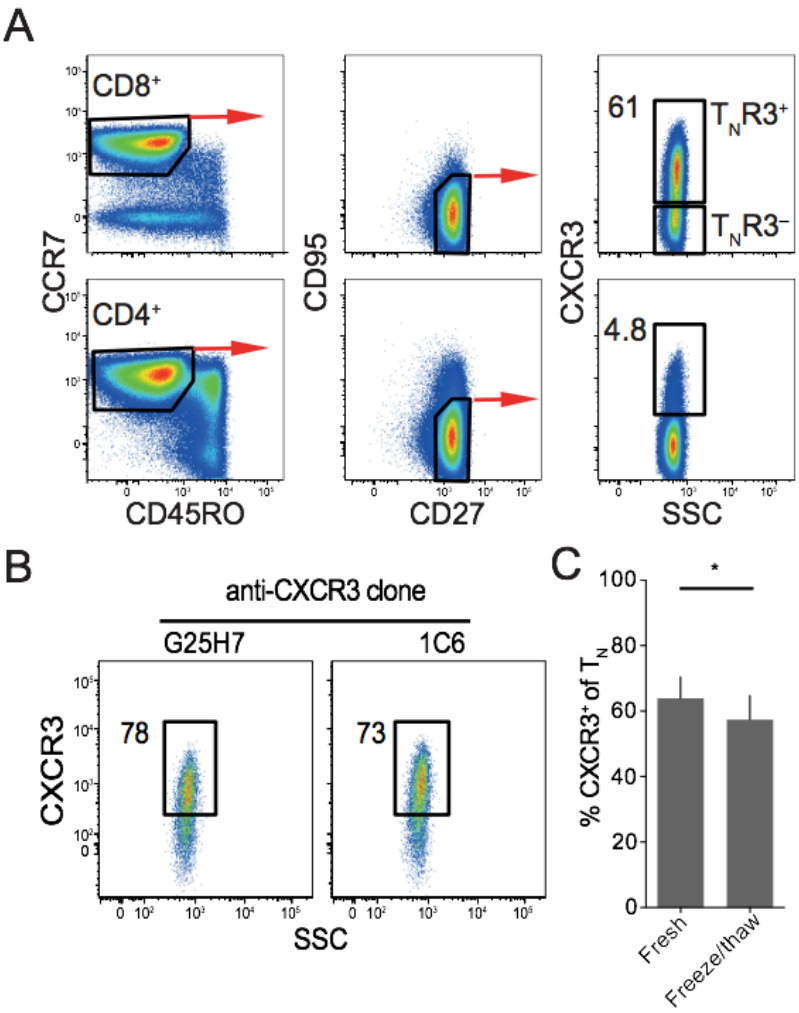


Figure 4: CXCR3 identifies two subsets of T_N-phenotype cells in the human body. A) Representative flow cytometry analysis of CXCR3 expression by $CD45RO^{-}CCR7^{+}CD27^{+}CD95^{-}$ T_N-phenotype $CD8^{+}$ and $CD4^{+}$ T cells. Numbers indicate the percentage of positive cells identified by the gates. **B)** Representative flow cytometry analysis of CXCR3 expression on $CD8^{+}$ T_N cells as detected by two clones of fluorescently-conjugated antibodies. Similar data were obtained from 4 more individuals. **C)** Mean±SEM percent CXCR3 expression by $CD8^{+}$ T_N cells identified in PBMCs that are fresh or after freezing/thawing in liquid nitrogen (n=5). *=P<0.05, paired t-test.

clearly distinguish $CXCR3^{-}$ and $CXCR3^{+}$ $CD8^{+}$ T_N cells (from here referred as T_NR3^{-} and T_NR3^{+}, respectively) (Fig 4A). This unprecedented heterogeneity within the T_N compartment allowed us to perform an in depth characterization of these two newly discovered subsets, which constitute the $CD8^{+}$ T_N pool showed in Fig 1. We confirmed CXCR3 expression on the surface of $CD8^{+}$ T_N cells with two different anti-CXCR3 antibody clones, obtaining similar frequencies of positive cells (Fig 4B). There is evidence that chemokine receptor expression is affected by cycles of freezing-thawing[132]; thus we compared CXCR3 expression on freshly isolated PBMCs and on cryopreserved PBMCs from the same donors. Results in Fig 4C show a reduction in CXCR3 expression on thawed samples, therefore we performed most of our assays using freshly isolated PBMCs.

We next aimed to confirm these differences in protein expression also at the mRNA level, and thus developed a FACS sorting panel in order to isolate T_NR3^{-} and T_NR3^{+} for additional studies. We thus flow-sorted T_NR3^{-} and T_NR3^{+} from healthy controls (HC) PBMCs and confirmed that *CXCR3* mRNA is detected in all donors in T_NR3^{+} cells while was lower or undetected in T_NR3^{-} cells (Fig 5A).

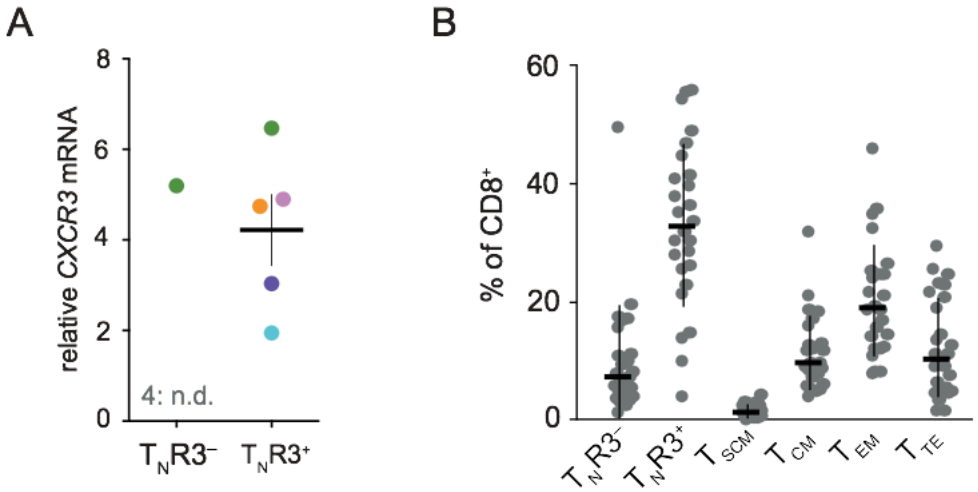


Figure 5: **CXCR3 mRNA quantification of T_N subsets and their relative abundance amongst total $CD8^+$T cells. A)** Mean±SEM expression of *CXCR3* relative to *B2M* mRNA in FACS-sorted T_NR3^- and T_NR3^+ cells (n=5). N.d.: not detected. Same color indicates same donor of provenience. **B)** Mean±SEM frequency of T cell subsets in the peripheral blood of healthy individuals (n=26). T_{SCM}, Stem Cell Memory: CD45RO $CCR7^+CD27^+CD95^+$; T_{CM}, Central Memory: $CD45RO^+CCR7^+$; T_{EM}, Effector Memory: $CD45RO^+CCR7^-$; T_{TE}, Terminal Effectors: $CD45RO^-CCR7^-$.

In order to define the relative abundance of T_NR3^- and T_NR3^+ amongst total $CD8^+$ T lymphocytes, we performed phenotypic analysis of peripheral blood (PB) from HCs via FACS. Fig 5B shows that T_NR3^+ are ~3-fold more abundant compared to T_NR3^- under physiological conditions. We then investigated the immunophenotypes of T_NR3^- and T_NR3^+ subsets. To this aim, we took advantage of a publicly available data set from Wong et al., who analyzed the surface proteome of $CD8^+$ T cells from several human tissues by CyTOF[124]. This included markers of differentiation, homing and activation/costimulation. As depicted in Fig 6A, both T_NR3^- and T_NR3^+

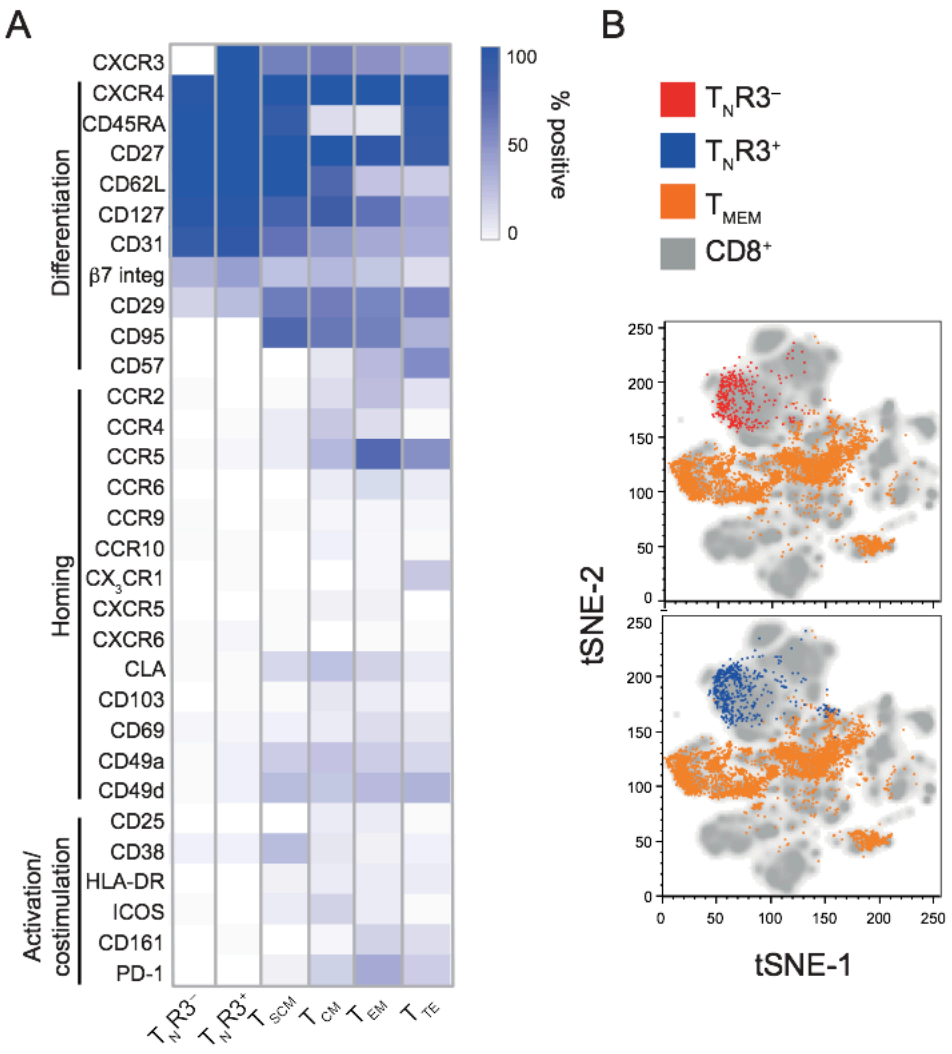


Figure 6: Phenotypic analysis of human T_N subsets. A) Heatmap of percent antigen expression by $CD8^+$ T cell subsets identified in the peripheral blood. Memory subsets were defined as in Fig 5B**. B)** t-SNE map representation of surface immunophenotypes of circulating $T_NR3^–$, T_NR3^+ and T_{MEM} cells (defined as $CD45RO^+$) on top of total $CD8^+$ T cells from the different tissues analyzed, as obtained with CyTOF. All markers listed in A, except CXCR3, were used for the t-SNE map.

displayed a T_N differentiation profile, characterized by the simultaneous expression of CD45RA, CD27, CD62L and CD127 in concomitance with the absence of CD95 and CD57. Analysis of homing and activation/costimulation molecules also revealed a clear T_N profile, with the lack of expression of tissue residency markers, integrins, chemokine receptors and activation markers (CD69, CD103, CD49a, CD49d, CCR5,

CCR6, CLA, CD38, HLA-DR, PD-1 and others) generally expressed by memory cells. Taken together, these observations revealed a shared phenotype between T_NR3^- and T_NR3^+ cells. In line with this observation, dimensionality reduction analysis performed with t-distributed stochastic neighbor embedding (tSNE) showed that T_NR3^- and T_NR3^+ have a similar profile, instead conventional $CD45RO^+$ memory T cells mapped in a clearly distinct region (Fig 6B).

4.1.1.2 Tissue homing and distribution of T_NR3^- and T_NR3^+

Under physiological conditions, human T_N cells are preferentially present in blood and lymphoid tissues[133]. Consistently with this observation, analysis of tissues from the same dataset of Fig 6A revealed that T_NR3^- and T_NR3^+ are not present in gut, skin and lung, while are present in tonsils and, with at a lesser extent, in spleen and liver (Fig 7A). Furthermore, we also investigated by FACS the presence of T_NR3^- and T_NR3^+ in lymph-nodes (LN) obtained from patients with head and neck cancer and in paired PB samples, and found comparable frequencies (Fig 7B). Surface markers analysis, performed as in Fig6A, displayed that T_NR3^- and T_NR3^+ in tonsil, spleen and liver possess an identical phenotype compared to PB, and that both subsets lacked the expression of tissue-residency markers CD69 and CD103. Moreover, t-SNE analysis performed on the same tissues displayed a similar spatial distribution of subsets compared to the same analysis on PB (Fig 7D). These data suggest that CXCR3 does not affect $CD8^+$ T_N cell trafficking.

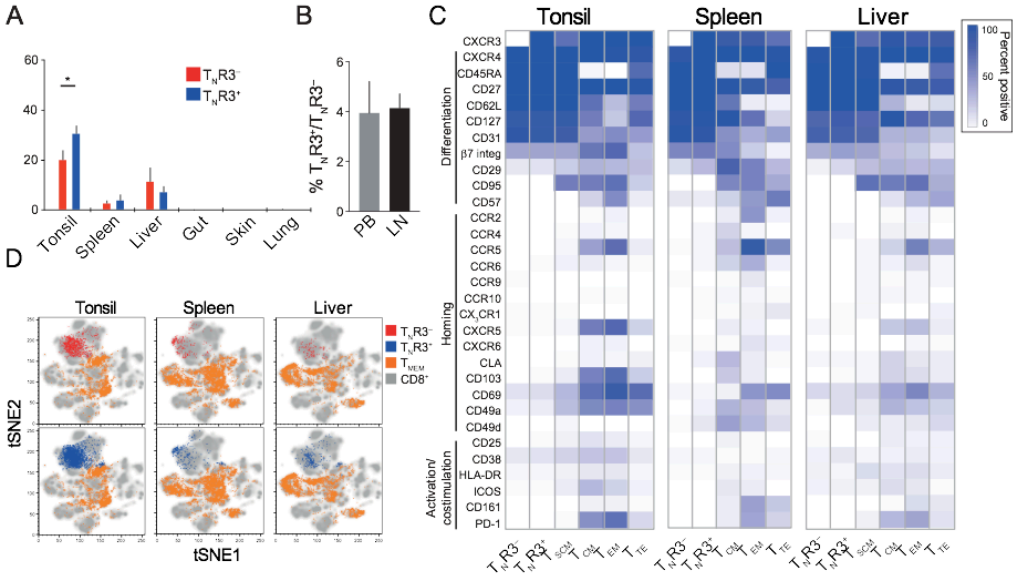


Figure 7: Tissue distribution of T_N subsets. A) Mean±SEM frequency of T_NR3^- and T_NR3^+ cells among total $CD8^+$ T cells isolated from human tonsils (n=5), spleen (n=3), liver (n=3), gut (n=6), skin (n=5), lung (n=4), as determined by CyTOF. *=P<0.05, Student's t-test. **B)** Percent ratio of T_NR3^+/T_NR3^- cells in paired LN and PB samples. Data are shown as mean ± SEM. **C)** Percent antigen expression by $CD8^+$ T cell subsets identified in the indicated tissues. Memory subsets were defined as in Figure 5B. **D)** tSNE map representation of surface immunophenotypes of T_NR3^- and T_NR3^+ cells along with T_{MEM} cells (defined as $CD45RO^+$) on top of total $CD8^+$ T cells from the indicated tissues. All markers listed in B except for CXCR3 were used for the tSNE map.

4.1.1.3 Abundance of T_NR3^- and T_NR3^+ during aging

Aging is characterized by a decline in the functionality of the immune system [134], in particular T cell responsiveness and pool composition are affected by advanced age[135]. Since T_N cell frequency is known to decrease with age, we investigated the balance of T_NR3^- and T_NR3^+ in the PB of a cohort of 1938 individuals with an age range from 19 to 105 years from a previously reported cohort[108,109]. Fig 8 displays that both T_NR3^- and T_NR3^+ decreased with age, but with different slope (Fig 8A and 8B). Frequency of T_NR3^+, in fact, resulted slightly increased with aging, thus

T_NR3^+ tend to become more abundant in PB during age progressing (Fig 8C).

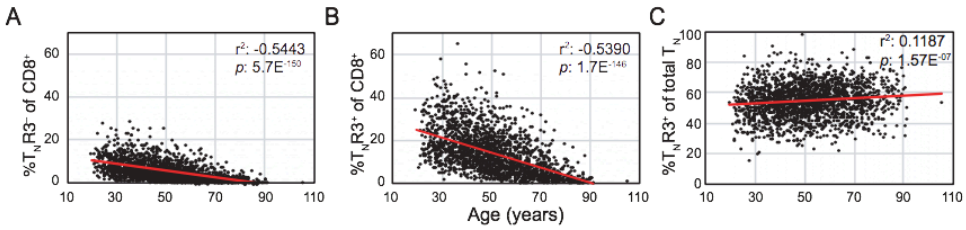


Figure 8: Abundance of T_NR3^- and T_NR3^+ during aging. A) Frequency of T_NR3^-cells on total $CD8^+$ in 1,938 individuals with different age (19-105 y.). Red line indicates linear regression. Effect size and p-value are indicated for each correlation. **B)** Frequency of T_NR3^+ on total $CD8^+$ in the same cohort in **A. C)** Frequency of T_NR3^+ on total T_N in the same cohort in **A** and **B**.

4.1.2 True naivety of T_NR3^+ cells

4.1.2.1 Replicative history of T_NR3^- and T_NR3^+

Previous work from Song and colleagues focused on a small subset of $CXCR3^+$ cells within the $CD4^+$ T_N compartment[136] (Fig 4A). Interestingly, these cells display functional responses and gene expression profile typical of memory cells, and resulted precursors of Th1 cells. Also recently work from Murata et al. identified a $CXCR3^+$ $CD8^+$ T_N-like population as young memory T cells[137]. Thus we performed a plethora of assays with the aim to investigate whether T_NR3+ cells might represent a subset of early differentiated $CD8^+$ memory T cells or a population of true naïve cells, and to deeply characterize the naïve vs memory features of T_NR3^- and T_NR3^+. First, we move to analyze the replicative history of T_NR3^- and T_NR3^+, by measuring T Cell Receptor Excision Circles content (TRECs). TRECs are

progressively diluted but not replicated upon cell division, thus providing the

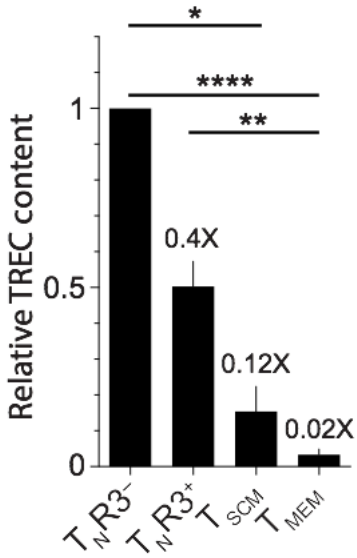


Figure 9: Replicative history of T_NR3^- and T_NR3^+. Mean±SEM copies of TREC as relative to T_NR3^- in FACS-purified T cell subsets isolated from the peripheral blood (n=10 for all subsets except for T_{SCM}, n=4). Numbers indicate the fold change of copies as relative to T_NR3^-. *=P<0.05, **=P<0.01, ****=P<0.0001, non-parametric ANOVA with Dunn's posttest.

possibility to effectively quantify cell divisions underwent by a specific cell population. Ex vivo assay for quantification of TRECs copies measured by quantitative real-time PCR[138] revealed that TRECs were ~ 2-fold more abundant in T_NR3^- compared to T_NR3^+, ~ 4-fold more abundant in T_NR3^- compared to T_{SCM} and ~ 25-fold more abundant in T_NR3^- compared to conventional $CD45RO^+$ memory T cells (Fig 9). According to data, T_NR3^+ underwent on average 1 additional cycle of cell division in vivo, at the population level, compared to T_NR3^-. This mild difference suggests a correlation between homeostatic proliferation and the expression of CXCR3.

4.1.2.2 Antigen specific clonal expansion of T_NR3^- and T_NR3^+

Naïve T cell compartment is characterized by the absence of clonally expanded cells, specific for exogenous antigens, but in some cases are present large number of precursors specific for self-antigens[139]. In particular, precursor frequency of Melan-A/MART-1-specifc T cells is at least 100-fold larger than that of T_N cells with other specificities[140]. The mechanism at the basis of this difference is poorly understood, but is known to originate in the thymus. Naivety of Melan-A/MART-1-specifc T cells, as revealed by TREC dilution and telomere length, is maintained in the periphery, and is thus conceivable with the concept of antigen ignorance[140].

According to these previous observations, T_NR3^- and T_NR3^+ does not display cells specific for immunodominant epitopes derived from CMV and Flu, while cells specific for the same epitopes are present in T_{SCM} and in T_{MEM} populations. Furthermore, cells specific for immunodominant epitope derived from Melan-A/MART-1 were present in both T_NR3^- and T_NR3^+ (Fig 10A and B). Recently, work from Pulko and colleagues identified in the $CD8^+$ T cell compartment a rare population of memory cell with a naïve phenotype (T_{MNP}), which are characterized by the expression of CD49d and high levels of CXCR3, and by the capability to produce IFN-γ after stimulation with mitogens (PMA and Ionomycin)[141]. Furthermore, T_{MNP} revealed to include cells specific for epitopes derived from persistent viruses, like CMV and EBV; on the contrary, they do not display cells specific for epitopes derived from acute viruses, like Flu. In contrast with these data, we observed Flu-specific cells $CD8^+$ T cells, but not CMV-specific ones within the T_{MNP} compartment, which is included in the T_NR3^+ cell population (~ 0,5% of the T_NR3^+ subset) (Fig 10A and B).

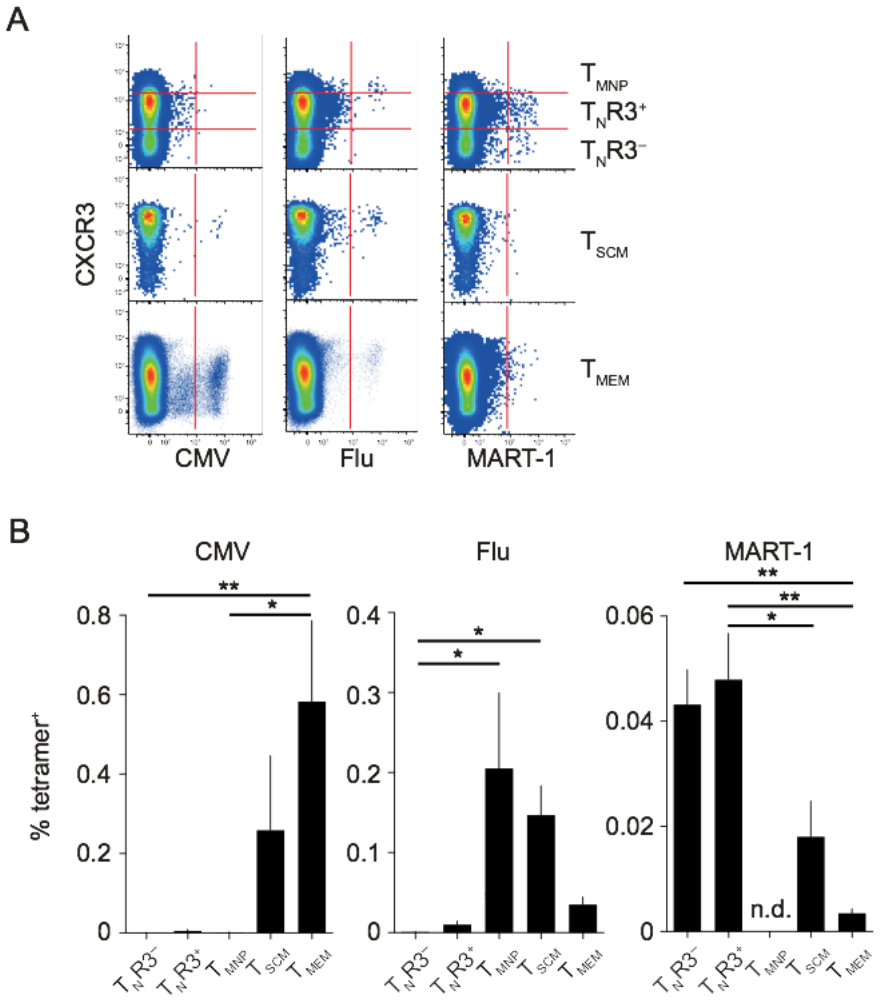


Figure 10: Antigen specific clonal expansion of T_NR3^- and T_NR3^+. A) Representative flow cytometry analysis of CMV, Flu and MART-1 HLA-A*02 tetramer binding in relation to CXCR3 expression in peripheral blood T cell subsets. Top row shows $CD45RO^-CCR7^+CD27^+CD95^-$ T_N-phenotype cells, further gated as T_NR3^-, T_NR3^+ and T_{MNP} cells. **B)** Mean±SEM percent of the data as in C (CMV, n=6; Flu, n=7; MART-1, n=8). N.d.= not detected due to low numbers. *=P<0.05 and **=P<0.01, non-parametric ANOVA with Dunn's posttest.

4.1.2.3 Ag-specific precursors compartmentalization in $CD8^+$ T_N subsets

Data shown in Fig 10A and B on MART-1 Ag-specific cells suggested that Ag-specific precursors are not preferentially compartmentalized in one determined subset of $CD8^+$ T_N cells. To confirm this hypothesis, we

investigated the presence of cells specific for exogenous Ags within T_NR3^- and T_NR3^+. We screened amplified libraries of T_NR3^- , T_NR3^+ and CD45RO$^+$ T_{MEM} with peptide-pulsed autologous APCs[112]. CD8$^+$ T cells specific for previously unencountered antigens, like HIV-1 and Zika virus, were detected in both T_NR3^- and T_NR3^+

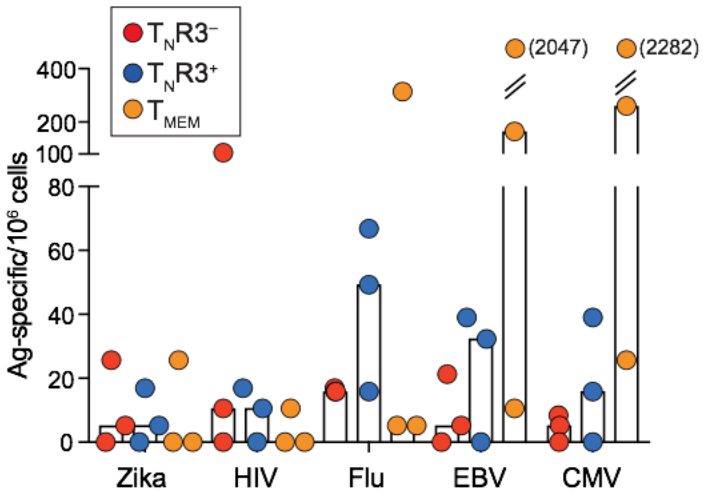


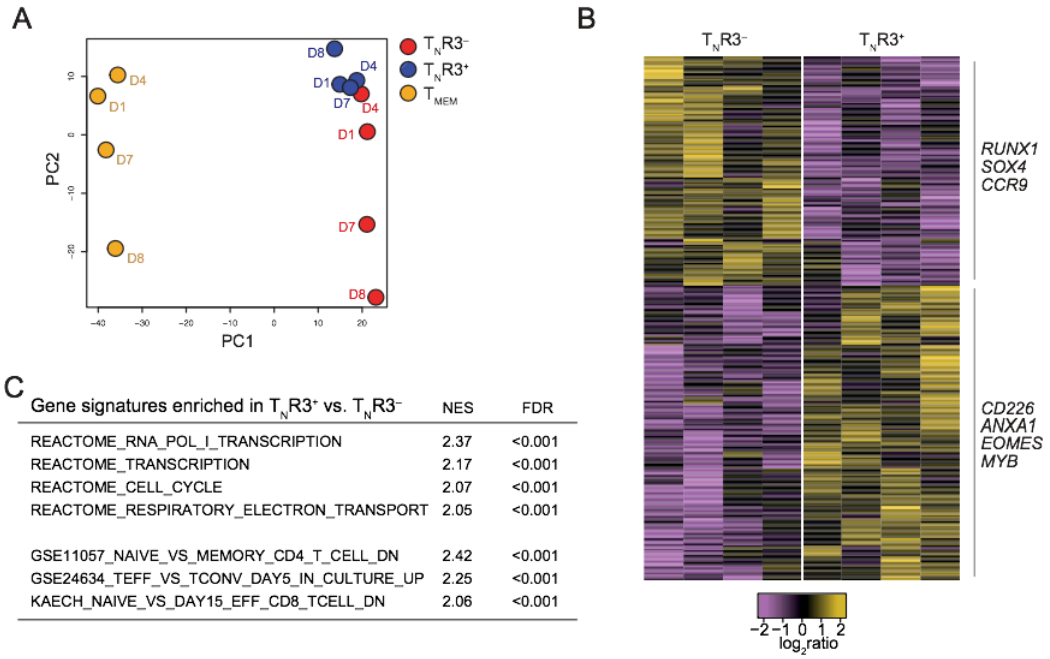
Figure 11: Ag-specific precursors compartmentalization in CD8$^+$ T_N subsets. Frequency of antigen-specific CD8$^+$ T cell precursors among T_NR3^-, T_NR3^+ and T_{MEM} cells. Bars indicate the median of the distribution while each dot represents a different donor (n=3).

compartments and with a similar extent. Only one donor constituted an exception, displaying high frequencies of HIV-1-specific only within T_NR3^- cell pool, probably indicating TCR cross-reactivity. Interestingly, Ag-specific cells for the more common viruses investigated in previous experiments (CMV, EBV and Flu) confirmed a similar trend even with this different experimental setting. Furthermore, CD8$^+$ T_{MEM} cells did not display cells specific for previously unencountered antigens, while on the other hand, Ag-specific cells for more prevalent viruses resulted abundant in T_{MEM} compartment (Fig 11).

4.1.3 T_NR3^+ are biased towards effector differentiation

4.1.3.1 T_NR3^+ cells display an effector-prone gene expression profile

After achieving information regarding phenotypical and functional characterization of T_N cell subsets we then move to assess relationship between T_NR3^- and T_NR3^+ and T_{MEM} at the gene expression level. In order to do this, we FACS-sorted matched T_NR3^- and T_NR3^+ and T_{MEM} from 4 HCs and subsequently performed RNA Microarray assay. To obtain immediate hints regarding similarity of subsets investigated we performed a principal component analysis (Fig 12A), which revealed that T_NR3^- and T_NR3^+ are predominantly distinct from $CD45RO^+$ T_{MEM} cells. Furthermore, T_NR3^- and T_NR3^+ resulted spatially very close in the plot, thus displaying a high degree of similarity in terms of gene expression profile. Paired analysis of the samples identified 345 differentially expressed genes (DEGs) with $p < 0.01$ between T_NR3^- and T_NR3^+ cells. As shown in the heat map depicted in Fig 12B, effector/memory-related transcripts like *ANXA1*, *EOMES* and *MYB*, and the costimulatory receptor *CD226*, which encodes DNAX accessory molecule-1 (DNAM-1) resulted upregulated in T_NR3^+ cells. Applying a less stringent cut-off ($p < 0.05$) was possible to identify 2567 DEGs, which included among others *BHLHE40*, a transcription factor related to effector differentiation, and *NT5E*, which encodes CD73, a surface enzyme involved in adenosine generation, both preferentially expressed by T_NR3^+ cells. Interestingly, T_NR3^- cells displayed overexpression of transcription factors associated with the regulation of thymocyte differentiation like *RUNX1*, *SOX4* and *IKZF1*.



Gene signatures enriched in T_NR3^+ vs. T_NR3^-	NES	FDR
REACTOME_RNA_POL_I_TRANSCRIPTION	2.37	<0.001
REACTOME_TRANSCRIPTION	2.17	<0.001
REACTOME_CELL_CYCLE	2.07	<0.001
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	2.05	<0.001
GSE11057_NAIVE_VS_MEMORY_CD4_T_CELL_DN	2.42	<0.001
GSE24634_TEFF_VS_TCONV_DAY5_IN_CULTURE_UP	2.25	<0.001
KAECH_NAIVE_VS_DAY15_EFF_CD8_TCELL_DN	2.06	<0.001

Figure 12: T_NR3^+ cells display an effector-prone gene expression profile. A) Principal component analysis based on the expression levels of genes with a coefficient of variation larger than the 90th percentile of the coefficients of variation in the entire dataset as obtained from microarray analysis (n=4/subset). **B)** Heatmap of DEGs between T_NR3^- and T_NR3^+ cells (n=4; P<0.01). The most relevant genes involved in immune functions are listed. **C)** Normalized enrichment score (NES) and false discovery rate (FDR) of the indicated gene signatures enriched in T_NR3^+ vs. T_NR3^- cells, as obtained by GSEA.

We next performed gene set enrichment assay (GSEA) in order to capture global transcriptional trends at the level of gene signatures rather than at the level of single genes. This analysis displayed that gene sets involved in mitochondrial activity like respiratory electron transport chain (REACTOME Database) and pathways of genes associated with effector and memory differentiation (Immunological Signature Database) were significantly enriched in T_NR3^+ versus T_NR3^- cells (FDR < 0.001; Fig 12C). Overall, these data suggested that T_NR3^+ cells are better poised to differentiate and acquire effector functionality compared with T_NR3^- cells.

4.1.3.2 T_NR3^+ cells possess increased cytokine production *in vitro*

We then moved to assess whether these observations regarding gene expression profile reflect on increased effector functions capability in T_NR3^+ cells compared to TNR3$^-$ cells. In order to do this, we FACS-sorted T_NR3^- and T_NR3^+ from HCs alongside with T_{MEM}, stimulated directly *ex vivo* with PMA and ionomycin for 6 hours and measured cytokine production by FACS. As depicted in Fig 13A, T_NR3^+ cells resulted capable to produce higher amount of TNF-α and IL-2 compared to the T_NR3^- counterpart. Increased IL-2 production might be related to the lower expression of IKZF1 in T_NR3^+, which is a negative regulator of IL-2

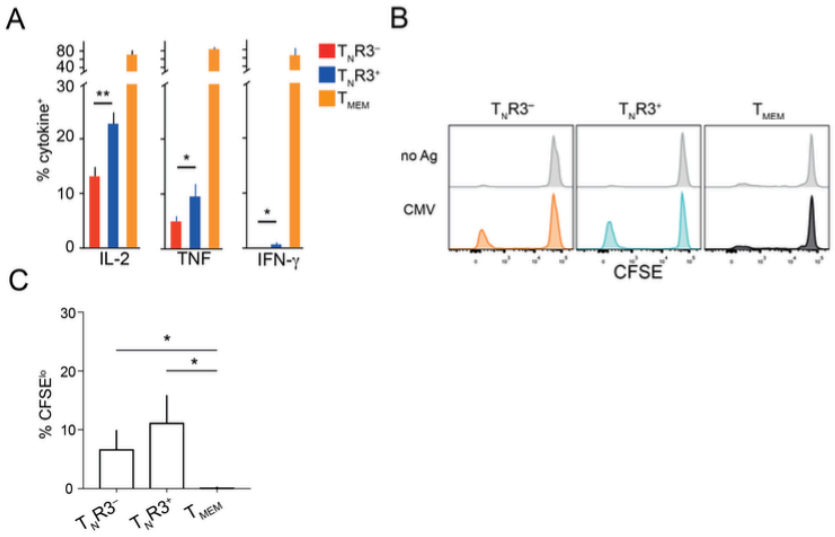


Figure 13: T_NR3+ cells possess increased cytokine production in vitro. **A) A)** Mean±SEM percent of FACS-purified subsets producing IL-2, TNF-α and IFN-γ following PMA/ionomycin stimulation for 6 hours (n=6 for T_NR3^- vs. T_NR3^+, n=2 for T_{MEM}). **B)** Representative CFSE dilution at day 10 of T cell subsets following HCMV lysate and CMV peptide pool stimulation. **C)** Mean±SEM percent of CMV-specific $CD8^+$ T cells diluting CFSE, obtained as in **B** (n=5). *=P<0.05, non-parametric ANOVA (Friedman test).

production in CD8⁺ T cells[142]. Furthermore, according to their naïve status, no IFN-γ production was detected in both T_NR3^- and T_NR3^+ cells, thus providing additional confirmation of the naivety of both subsets. We then investigated effector functions capability of T_N subsets at the Ag-specific level. To do this, we FACS-sorted T_NR3^-, T_NR3^+ and T_{MEM} from CMV-seronegative donors and cultured for 10 days with autologous monocytes pulsed with human CMV lysate and a CMV peptide pool. CMV-specific response was detected in both T_NR3^- and T_NR3^+ but not in the T_{MEM} pool, as assessed by CFSE dilution (Fig 13B and C). Moreover, on day 10, cells were stimulated with PMA and ionomycin to measure cytokine production,

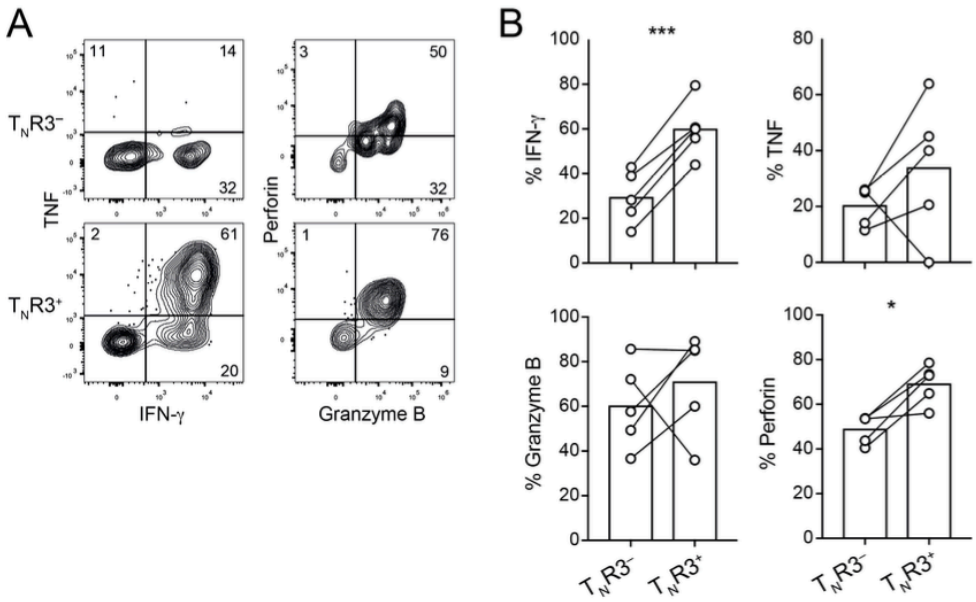


Figure 14: Ag-specific T_NR3^+ cells possess increased cytokine production in vitro. A) Representative flow cytometry analysis of effector molecules production following PMA/ionomycin stimulation by CFSE- diluted, CMV-specific T_NR3^- and T_NR3^+ CD8⁺ T cells, obtained as in **10B**. Numbers indicate the percentage of cells identified by the gates. **B)** Mean summary of the data as in **A**. Each dot represents a different donor. * P<0.05 and *** P<0.001, paired t-test.

and an increased expression of IFN-γ, perforin and TNF-α was observed in Ag-specific CFSE-diluted T_NR3^+ cells in comparison with correspondent T_NR3^- (Fig 14A and B).

4.1.3.3 T_NR3^+ cells possess increased cytokine production *in vivo*

To extend our in vitro observations on human CMV-specific precursors to a physiological setting in vivo, we investigated whether phenotypically similar T_N cell subsets could be identified in the mouse. Indeed, we could observe that a subpopulation of murine $CD44^{lo}CD62L^{hi}$ $CD8^+$ T_N cells expresses CXCR3, although with a different frequency compared to human T_N cells (~10% vs. ~60%; Fig 15A and Fig 4A, respectively). Next, we FACS-purified murine T_NR3^-

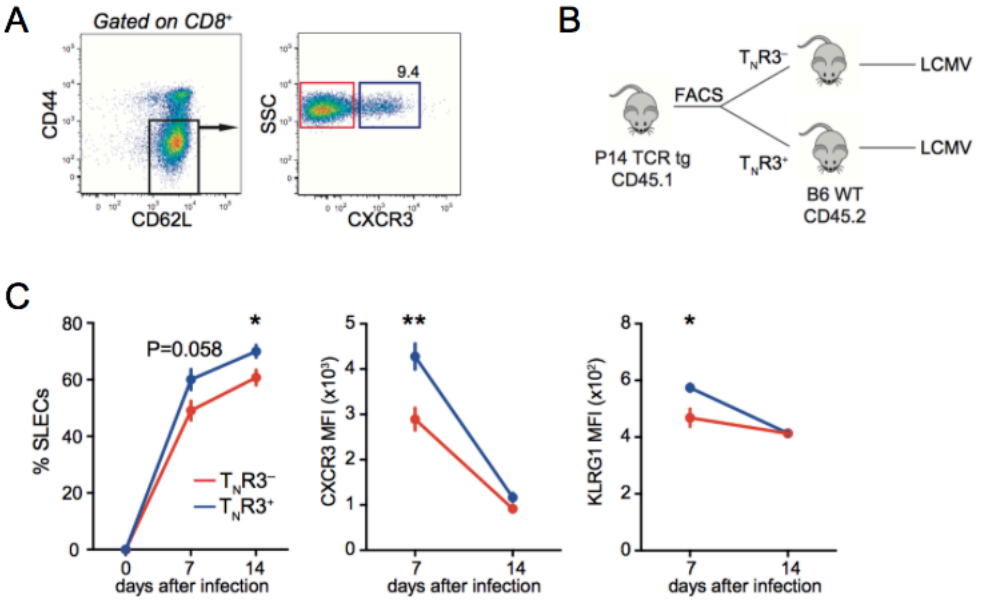


Figure 15: Ag-specific T_NR3^+ cells possess increased cytokine production in vivo. A) Representative gating strategy for the identification of murine T_NR3^- and T_NR3^+ cells in LCMV P14 TCR transgenic mice. **B)** Schematic representation of the murine adoptive transfer experiment. **C)** Mean±SEM frequencies (%) of spleen $KLRG1^+$ $IL\text{-}7Ra^-$ SLECs among adoptively transferred P14-specific T_NR3^- and T_NR3^+ $CD8^+$ T cells (left panel) and CXCR3 and KLRG1 MFI in SLECs (center and right panel, respectively) at different days post LCMV infection (n=5 mice/group). *=P<0.05 and **=P<0.01, unpaired parametric t-test.

and T_NR3^+ cells from unimmunized, LCMV P14 TCR-transgenic mice, and adoptively-transferred them into wild type hosts that were subsequently infected with the LCMV Armstrong strain (causing an acute infection; Fig 15B). By analyzing the progeny of these cells, distinguished from the endogenous response by specific congenic markers, we found that murine T_NR3^+ cells generated a higher frequency of KLRG1$^+$ IL-7Ra$^-$ SLECs at day 7 and day 14 after infection. Moreover, SLECs deriving from T_NR3^+ cells featured higher levels of CXCR3 and KLRG1 protein expression at day 7 post infection than their T_NR3^- counterparts (Fig 15C). We conclude that also murine T_NR3^+ cells are biased towards effector differentiation following antigen-specific stimulation in vivo.

4.1.4 T_NR3^- and T_NR3^+ cells express qualitatively distinct TCRs

We then decided to investigate whether this bias in terms of effector function and differentiation of T_NR3^+ compared to T_NR3^- could be related to different intrinsic features of TCR reactivity. Tonic TCR signaling driven by self-p:MHC-I complexes is fundamental for T_N homeostasis and maintenance (see Introduction). In mice, increased TCR reactivity against self-Ags correlates with the expression of CD5 and indicates the ability of T_N to respond to foreign antigens[70–72]. Thus, we initially investigated levels of expression of CD5 on the surface of T_NR3^- and T_NR3^+. As depicted in Fig 16A, we found that human T_N subsets uniformly express CD5 on their surfaces, suggesting that CD5 may not identify differential T cell reactivity of human T_N cells. Alternatively, physicochemical properties of TRB CDR3 sequences, as determined by TCR sequencing, can indirectly inform on the strength of interaction between the TCR and p:MHC-I complexes[73,116,143–147]. We thus performed TCR deep sequencing on

FACS-sorted T_NR3-, T_NR3+ and T_{MEM} cells in order to investigate their respective TRB repertoires. First, we performed an analysis of TRBV-TRBJ

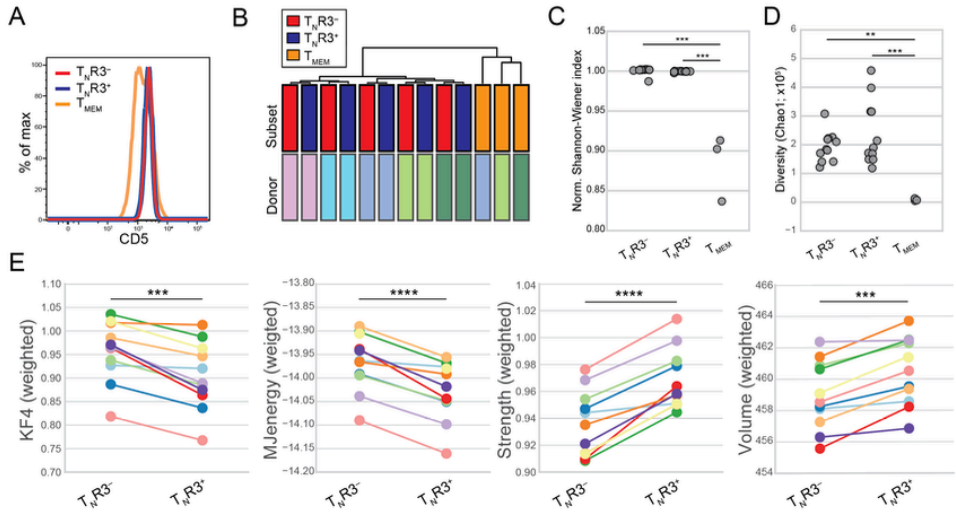


Figure 16 : T_NR3^- and T_NR3^+ cells express qualitatively distinct TCRs. **A**) Representative flow cytometric analysis of surface CD5 on human $T_NR3^–$, T_NR3^+ and T_{MEM} cells. Similar data were obtained from 3 more individuals. **B)** Cluster analysis of TRBV-TRBJ usage, as obtained by TCRseq, in sorted $T_NR3^–$ and T_NR3^+ cells (both n=5). Data are from a subset of individuals from a single sequencing experiment where also T_{MEM} cells (n=3) were analyzed. **C)** Normalized Shannon-Wiener diversity index and **D)** Chao1 diversity index of TCRseq data (calculated for 5,000 unique UMI-labeled TRB CDR3 molecules; n=11 from two independent experiments). ***= P<0.001, Tukey's range test. **E)** Averaged (weighted per clonal size) Kidera Factor 4 (KF4), MJ energy, Strength and Volume VDJtools physico-chemical characteristics of 5 amino acids in the middle of CDR3 sequences in $T_NR3^–$ and T_NR3^+ cells TRB repertoires, obtained as in D. *** 0.001>P>0.0001, **** P<0.0001, paired t-test.

usage on individual HCs, and found that T_NR3^- and T_NR3^+ display highly similar patterns according to Jenson-Shannon divergence analysis (Fig 16B), indicating that they are highly related. In contrast, T_{MEM} cells revealed distinct patterns of TRBV-TRBJ. According with their naïve

status, T_NR3^- and T_NR3^+ showed high and very similar diversity of TRB sequences, as displayed in Fig 16C and D by normalized Shannon-Wiener and Chao1 metrics calculated for 5,000 randomly chosen TCR beta cDNA molecules. As expected on the basis of their clonal expansion, diversity was much lower in T_{MEM} cells. We then moved to investigate the averaged physicochemical properties of the central region of each TRB CDR3 sequence, composed by 5 aa residues, which is mainly responsible for the contact with p:MHC-I complexes. As shown in Fig 16E, analysis by VDJtools[115] displayed that, at the population level, T_NR3^+ possess augmented hydrophobicity, as assessed by lower Kidera factor 4 (KF4), lower energy[117,144] and higher strength and volume parameters compared to the T_NR3^- cell pool. Taken together, these parameters suggest that T_NR3^+ are characterized by TCRs with augmented intrinsic affinities for cognate Ags (being that either self or non-self) compared to T_NR3^- cells.

4.1.5 T_NR3^+ cells are transcriptionally equivalent in human and mice

To corroborate findings of possible increased reactivity of T_NR3^+ vs. T_NR3^- cells with a defined model, we turned to analyze the murine polyclonal $CD44^{lo}$ $CD62L^{hi}$ $CD8^+$ T_N compartment. Murine T_NR3^+ T cells are contained in the $CD5^{hi}$ fraction (Figure 17A and B), and were previously

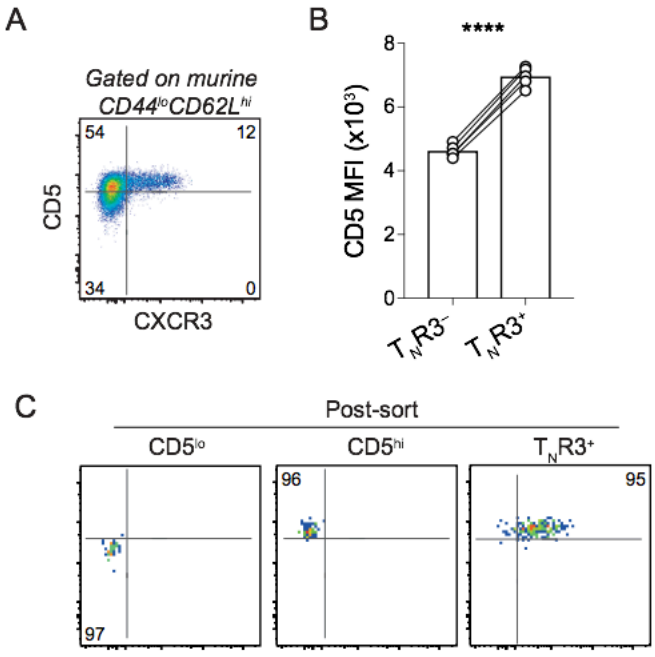


Figure 17 : Characterization of murine T_N subsets. A) Representative flow cytometry analysis of CXCR3 and CD5 expression by murine $CD44^{lo}CD62L^{hi}$ $CD8^+$ T_N cells. **B)** Mean fluorescence intensity (MFI) of CD5, as determined by flow cytometry, by murine T_NR3^- and T_NR3^+ cells (gated as $CD44^{lo}CD62L^{hi}CXCR3^-$ and $CD44^{lo}CD62L^{hi}CXCR3^+$, respectively; n=5; P<0.0001, paired t-test). **C)** Post-sort purity, as obtained by FACS, of $CD8^+$ T_N subsets identified by differential CD5 and CXCR3 expression.

shown to respond more vigorously to foreign antigens compared to $CD5^{lo}$ and $CD5^{hi}$ T_N cells. In order to obtain an all-encompassing characterization of T_N heterogeneity in murine compartment, we FACS-sorted $CD5^{lo}CXCR3^-$ (referred to as $CD5^{lo}$), $CD5^{hi}CXCR3^-$ ($CD5^{hi}$) and $CD5^{hi}CXCR3^+$ (T_NR3^+) from the $CD44^{lo}CD62L^{hi}$ T_N cell pool (Fig 17C) along with $CD44^{hi}$ conventional memory T cells, and defined their transcriptional profile using RNA sequencing (RNAseq). As depicted in Fig 18A, according to biological coefficient of variation, all murine $CD8^+$ T_N subsets resulted clearly separated from conventional T_{MEM} cells. Nevertheless, gene expression analysis identified 636 DEGs (FDR <

0.001) among the murine T_N subsets. Besides *Cxcr3* and *Cd5*, murine T_NR3^+ had increased expression of effector T cell-associated genes including *Tbx21*, *Ccl5, Irf8, Hopx, Junb, Fos* and *Jun* compared to both $CD5^{lo}$ and $CD5^{hi}$ T_N cells, among others (Fig 18B) while, with less stringent criteria (FDR<0.05), decreased expression of naïve T cell-associated genes, including *Lef1* and *Ccr7*. Importantly, genes previously found differentially expressed between human T_N subsets such as *Ccr9*, *Eomes*, *Nt5e* and *Myb* or *Sox4* and *Ikzf1* were also found differentially expressed in murine T_N subsets. According to these data, gene signatures composed by the genes upregulated in murine T_NR3^+ vs. $CD5^{hi}$ and in murine T_NR3^+ vs. $CD5^{lo}$, containing 37 and 221 genes, respectively (FDR < 0.001), resulted preferentially enriched by GSEA in human T_NR3^+ compared to T_NR3^- cells. These data indicate that murine T_NR3^+ with enhanced effector T cell differentiation potential are transcriptionally equivalent to human T_NR3^+ counterpart (Fig 18C).

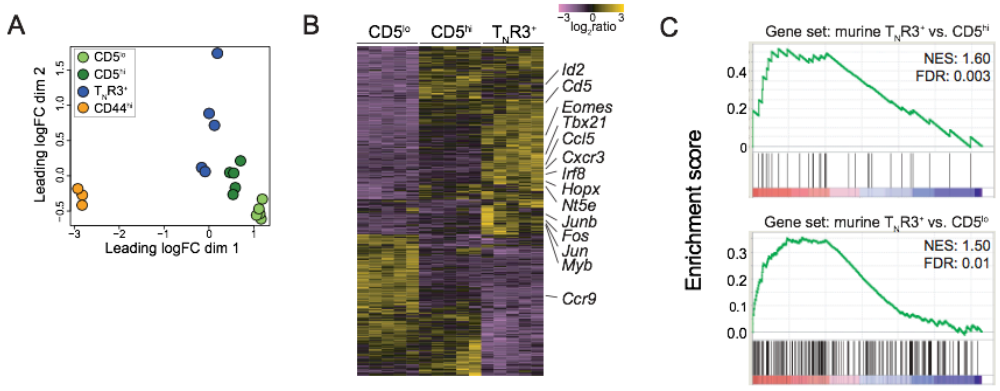


Figure 18 : Gene expression profile of murine T_N subsets. A) Biological coefficient of variation (BCV) plot of RNAseq profiles of murine $CD8^+$ T_N subsets isolated as in **15C** (n=5/subset). Bulk memory $CD8^+$ T cells are $CD44^{hi}$ (n=3). **B)** Heatmap of differentially expressed genes (FDR<0.0001) among subsets sorted as in **15C**. The most relevant genes involved in immune functions are listed. **C)** GSEA plots of murine T_NR3^+ vs. $CD5^{hi}$ (top) and vs. $CD5^{lo}$ (bottom) gene sets significantly enriched in human T_NR3^+ cells. NES: normalized enrichment score.

4.2 Results part 2

4.2.1 Heterogeneity of the human memory CD8⁺ T cell pool as revealed by high dimensional single cell analysis

4.2.1.1 Single-cell RNA sequencing of the human CD8⁺ T cell memory pool

We initially used single-cell RNA sequencing (scRNAseq; 10X Genomics platform) to characterize the full spectrum of human CD8⁺ memory T cells in peripheral blood (PB). A total of 31,640 cells were isolated for this purpose via fluorescence-activated cell sorting (FACS) based on the expression of CD95 which identifies a vast majority of all memory T cells in humans[25] (Fig 19A). Bioinformatic analysis of gene expression mapped

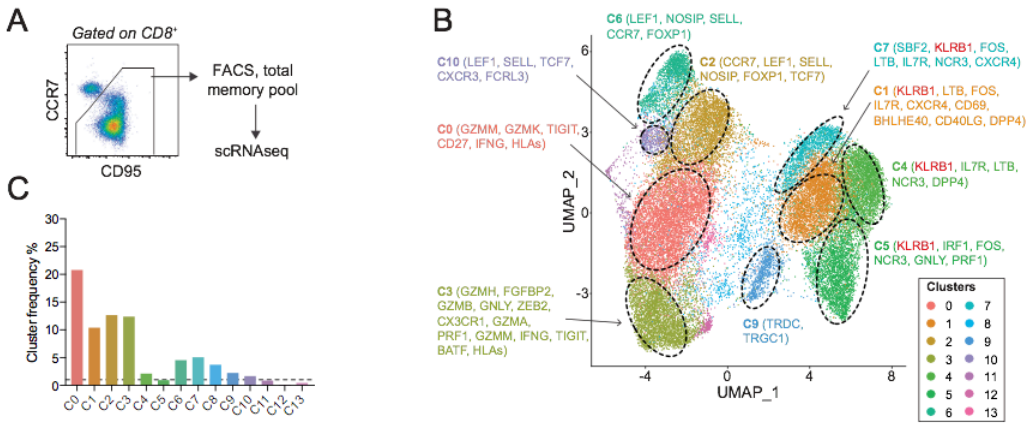


Figure 19 : Clusters identification in human memory T cell pool. A) Schematic representation of the sorting strategy of total memory CD95⁺ CD8⁺ T cells for scRNAseq processing. **B)** CD8⁺ memory T cells were sorted from four healthy donors' peripheral blood and processed for 10X Genomics scRNAseq. The UMAP plot shows the distribution of ~32,000 cells in 13 clusters. Selected genes shown are differentially expressed compared to all the other clusters. **C)** Histogram plot showing the frequency of each cluster derived from the scRNAseq. Dashed line sets a threshold at 1%.

in two dimensions via Uniform Manifold Approximation and Projection (UMAP)[148] (Fig 19B), a dimensionality reduction tool, which also takes in

consideration the relationship between cluster's similarity and spatial distance in the plot, identified 14 distinct clusters (denoted individually as C). As depicted in Fig 19B, were uniformly distant from the other cell populations and expressed high levels of *KLRB1*, which encodes CD161, and *IL7R*, which encodes the interleukin (IL)-7 receptor (IL-7R), also known as CD127. These clusters were therefore derived from mucosal associated invariant T (MAIT) cells[149]. An intermediate cluster, C9, which comprised less than 2% of all cells (Fig 19B) overexpressed *TRDC* and *TRGC1*, which encode the constant regions of the T cell receptor (TCR) δ and γ chains, respectively, suggesting the expression of $\gamma\delta$ rather than $\alpha\beta$ TCRs. In order to get information regarding the relative abundance of sub-populations, we plotted frequencies of clusters on bar graph (Fig 19C). We established a cut-off threshold of 1%, in order to exclude those really small populations which might not have a significant biological function.

4.2.1.2 Transcriptional heterogeneity in the CD8$^+$ memory T cell compartment consists of 5 distinct clusters

scRNAseq further identified seven different clusters related to conventional memory T cells (Fig 19B), the most abundant of which were C0, C2, C3, C6, and C10 (Fig 17C). C2, C6, and C10 expressed genes associated with long-lived memory T cells, including *LEF1*, *SELL*, which encodes L-selectin (CD62L), *CCR7*, and *FOXP1* (Fig 19B and 20A). Among these clusters, C10 lacked *CCR7* and expressed relatively high levels of Fc receptor-like protein 3 (*FCRL3*). Overall, these data suggest

transcriptional heterogeneity in the early differentiated memory T cell pool (Fig 19B and 20A).

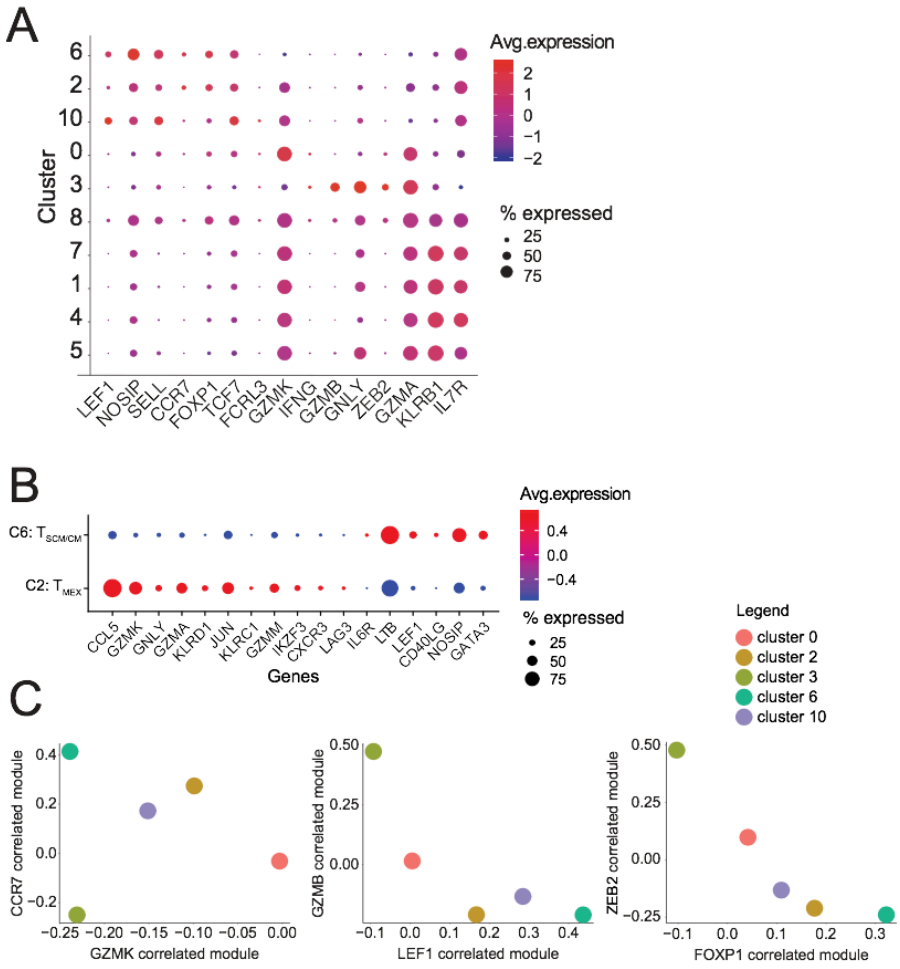


Figure 20 : Features of identified memory T cell clusters. A) and **B)** Balloon plot showing the average expression of selected genes and the frequency of expression within each cluster from **Fig 17B** in a comparison between (**A**) all the clusters or (**B**) C6 versus C2. **C)** Anchor gene analysis of the dataset in **Fig 17B** CCR7, GZMK, GZMB, LEF1, ZEB2 and FOXP1 were selected as anchor genes to obtain a list of correlated genes on the basis of Pearson correlation coefficient. This generated six gene correlated modules. The three graphs show the distribution of C0, C2, C3, C6 and C10 on the axis according to their correlation scores.

In contrast, the highly abundant C0 overexpressed multiple effector transcripts, including *GZMK* and *GZMM*, which encode serine proteases termed granzymes, *IFNG*, which encodes interferon (IFN)-γ, the chemokine (C-C motif) ligands *CCL4* and *CCL5*, and genes encoding

human leukocyte antigen (HLA) class II molecules, consistent with the identification of effector memory T (T_{EM}) cells (Fig 19B and 20A).

C3 displayed a gene expression profile reminiscent of terminal effector T (T_{TE}) cells, featuring high levels of *GZMB*, *GNLY, NKG7*, *ZEB2*, and *GZMA* (Fig 19B and 20A). A lack of signature transcripts precluded the identification of C8 based on current knowledge of the T cell differentiation pathway. C2, C6 and the less abundant C10 express genes associated with long-lived memory T cells, but with some peculiar differences. Both C2 and C6 are characterized by the expression of *CCR7*, *SELL*, *LEF1* and *TCF7* (which encodes for TCF1), while CCR7 expression was absent in C10, which instead display increased expression of *FCRL3*, which in $CD4^+$ T cells identifies a population of memory Tregs[150]. Taken together these preliminary observations on scRNAseq data, highlighted transcriptional heterogeneity within the $CD8^+$ early memory pool.

To explore the heterogeneity of the long-lived memory T cell pool, we focused on C2 and C6, because C10 comprised only ~1% of all sorted cells (Fig 19C). We identified 160 differentially expressed genes (DEGs) between C2 and C6. C2 expressed higher levels of effector molecules, including *CCL5*, *GZMK*, *GNLY*, *GZMA*, *JUN*, *GZMM*, *HOPX*, *IKZF3*, *RUNX3*, and *PRF1*, which encodes perforin, whereas C6 expressed higher levels of *IL6R*, *LTB*, *LEF1*, *NOSIP*, *GATA3*, and *SELL* (Fig 20B). We then used anchor genes selected from the most prominent DEGs to compute transcriptional modules associated with memory differentiation (correlated with *CCR7* and *LEF1*), quiescence (correlated with *FOXP1*), or cytotoxicity and terminal effector differentiation (correlated with *GZMK*/*GZMB* and *ZEB2*, respectively) among the five conventional memory T cell clusters (C0, C2, C3, C6, and C10). Using this approach,

we found that C2 and C10 were similar, exhibiting intermediate memory and effector scores, whereas C6 was skewed toward a high memory score and C0 and C3 were skewed toward high effector scores (Fig 20C). scRNAseq analysis therefore identified four major subsets of conventional memory $CD8^+$ T cells, namely early memory *CCR7*$^+$ *GZMK*$^-$ (C6) or *CCR7*$^+$ *GZMK*$^+$ (C2), T_{EM}-like *CCR7*$^-$ *GZMK*^{high} (C0), and T_{TE}-like *GZMB*$^+$ (C3).

4.2.1.3 Flow Cytometry characterization of memory T cell heterogeneity

To confirm scRNAseq data at the protein level, we designed a high-dimensional flow cytometry panel based on the cluster signature markers CCR7, LEF1, CD161, GZMB, and GZMK. This panel was also equipped to detect memory and effector differentiation markers (CD27, CD28, CD45RO, CD127, and T-bet), activation markers (CD38 and HLA-DR), inhibitory receptors [PD-1 and T cell immunoreceptor with Ig and ITIM domains (TIGIT)], and markers of tissue residency (CD69 and CD103). In line with the scRNAseq data, UMAP analysis of expressed proteins revealed that $CD161^{bright}$ MAIT cells were largely distinct from other $CD95^+$ memory T cells (Fig. 21A). The $CCR7^+$ $GMZK^-$ subset expressed relatively high levels of the memory markers LEF1, CD27, CD28, and CD127 and lacked effector molecules, activation markers, and inhibitory receptors, whereas the $CCR7^+$ $GMZK^+$ subset expressed intermediate levels of LEF1 alongside PD-1 and TIGIT, which were not detected in the scRNAseq analysis, together with relatively high levels of CD27, CD28, and CD127 (Fig. 21A). These subsets displayed variable expression of CD45RO (Fig. 21A). Among the effector subsets defined at the transcriptional level,

CCR7⁻ GZMK^{high} cells expressed cytolytic and effector molecules alongside relatively high levels of CD69, were LEF1^{low}, and lacked CD127 and the residency marker CD103, whereas GZMB⁺ cells also expressed granulysin (GNLY) and relatively high levels of T-bet (Fig. 21A). As expected, CD69⁺ CD103⁺ cells were only detected in tissues (Fig. 21A). A survey of different tissue sites revealed that CCR7⁺ GZMK⁻ PD-1⁻ TIGIT⁻ and CCR7⁺ GZMK⁺ PD-1⁺ TIGIT⁺ cells were relatively abundant in PB, lymph nodes (LNs), and bone marrow (BM), whereas CCR7⁻ GZMK^{high} cells were ubiquitous, and GZMB⁺ cells predominated in PB and lung tissue (Fig. 21B).

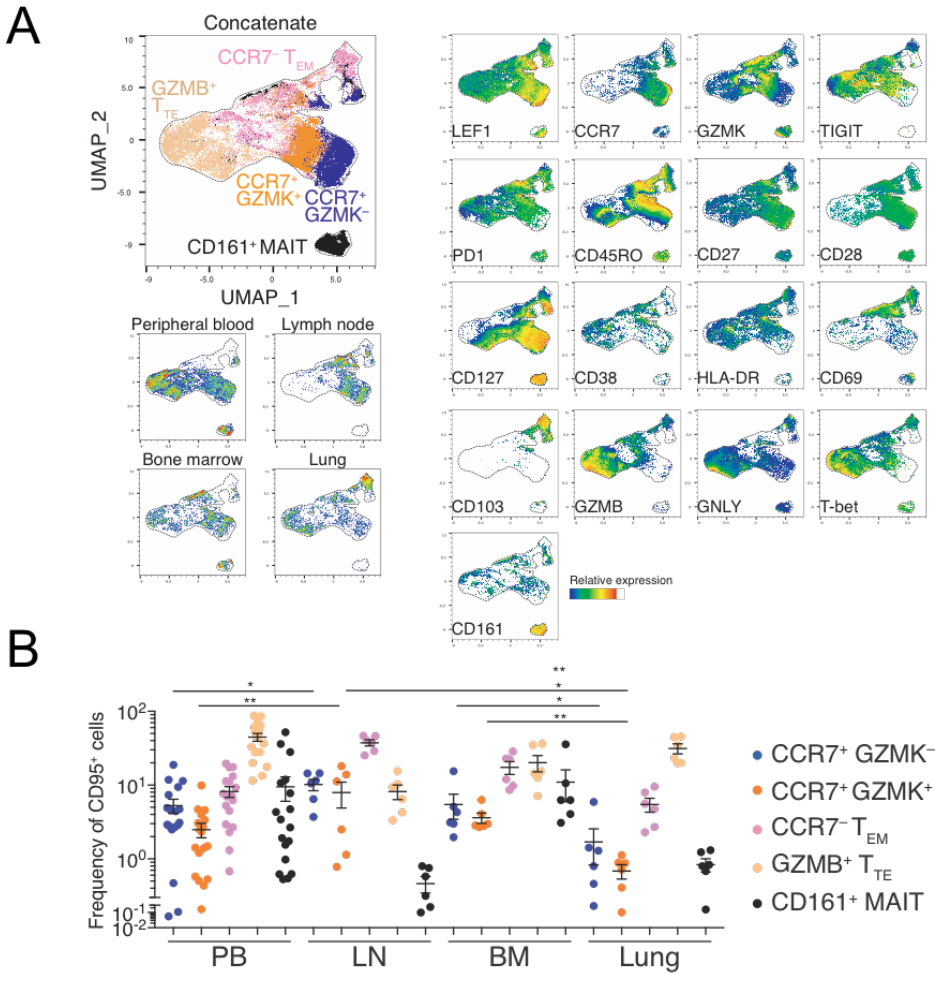


Figure 21 : Characterization of memory T cell clusters by flow cytometry. A) UMAP visualization of 20 parameters flow cytometry. The left graph displays UMAP clustering of the different concatenated(top) and individual (bottom) tissues. The top left graph displays an overlay with $CCR7^{+}GZMK^{-}$, $CCR7^{+}GZMK^{+}$, $CCR7^{-}$ T_{EM}, $CD161^{+}$ MAIT and $GZMB^{+}$ T_{TE}. The right graphs show the relative expression of each marker by positive cells for a specific marker on the UMAP (e.g. CD45RO expression by CD45RO positive cells). **B)** Frequency of the $CCR7^{+}GZMK^{-}$, $CCR7^{+}GZMK^{+}$, $CCR7^{-}T_{EM}$, $CD161^{+}$ MAIT and $GZMB^{+}$ T_{TE} subsets within the analyzed organs. Data are shown as Mean ± SEM. Statistics were calculated only for the $GZMK^{-}$ and $GZMK^{+}$ populations. *P < 0.05, **P <0.01 (Unpaired t test and Mann-Whitney test).

4.2.2 Conventional T_{SCM} and T_{CM} constitute the same pool of memory cells following removal of $GZMK^{+}$ cells with exhaustion traits

4.2.2.1 Exhausted-like $CD8^{+}$ memory T cell progenitors express GZMK, PD-1 and TIGIT

Heterogeneity in the long-lived memory T cell pool became apparent with the identification of T_{SCM} cells[25]. These cells exhibit a $CCR7^{+}$ $CD45RO^{-}$ $CD95^{+}$ phenotype, in contrast to T_{CM} cells, which exhibit a $CCR7^{+}$ $CD45RO^{+}$ $CD95^{+}$ phenotype. However, our scRNAseq-guided flow cytometric analyses demonstrated that the $CCR7^{+}$ $GZMK^{-}$ $PD\text{-}1^{-}$ $TIGIT^{-}$ and $CCR7^{+}$ $GZMK^{+}$ $PD\text{-}1^{+}$ $TIGIT^{+}$ subsets could not be distinguished via the expression of CD45RO (Fig 21A). To place these findings in context, we investigated the expression of PD-1 and TIGIT among classically defined T_{SCM} and T_{CM} cells. We found that 9.1 ± 1.3% of T_{SCM} cells and 22.1 ± 2.3% of T_{CM} cells (mean ± SEM) expressed both PD-1 and TIGIT (Fig 22A and B). On the basis of these results, we hypothesized that differential inclusion of the transcriptionally distinct $CCR7^{+}$ $GZMK^{+}$ $PD\text{-}1^{+}$

TIGIT$^+$ subset could explain some of the previously reported differences between T_{SCM} and T_{CM} cells.

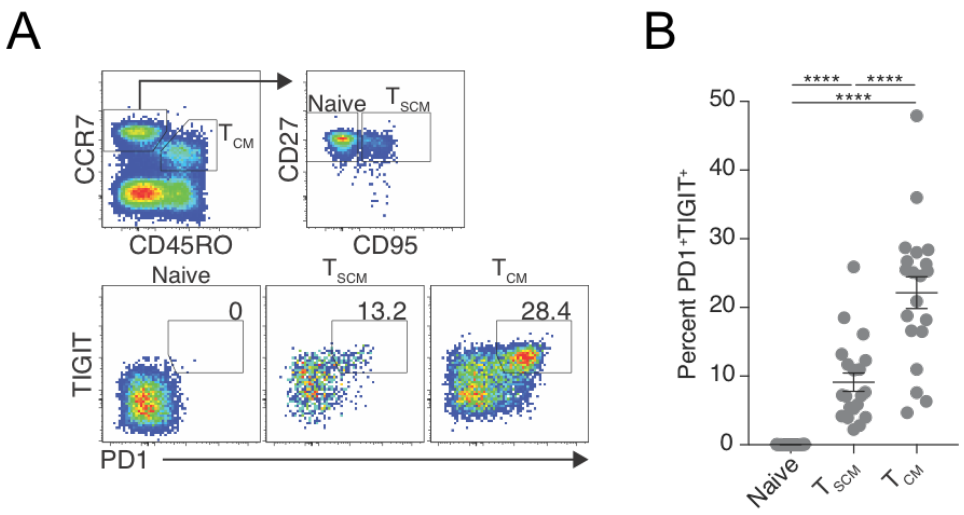


Figure 22 : Early differentiated memory pool displays a GZMK$^+$ contamination. A) Gating strategy showing the contamination of PD-1$^+$TIGIT$^+$ cells within classically defined Naïve (CCR7$^+$CD45RO$^-$CD28$^+$CD95$^-$), T_{SCM} (CCR7$^+$CD45RO$^-$CD28$^+$CD95$^+$) and T_{CM} (CCR7$^+$CD45RO$^+$) CD8$^+$ T cell subsets on total PBMCs of a representative donor. **B)** Frequency of PD-1$^+$TIGIT$^+$ cells within the classically defined Naïve, T_{SCM} and T_{CM} subsets (n=7, total PBMCs). Data are shown as median. **P < 0.01 (RM one-way ANOVA).

We confirmed by manual gating of the flow cytometry data that PD-1 and TIGIT were preferentially expressed by CCR7$^+$ GZMK$^+$ cells (not shown), thereby justifying the use of these surface markers to viably separate this subset of cells. We next analyzed the transcriptomes of T_{SCM} and T_{CM} cells after depletion of the CCR7$^+$ PD-1$^+$ TIGIT$^+$ (GZMK$^+$) population (hereafter termed progenitor exhausted-like, or T_{PEX}). In line with our hypothesis, such depleted subsets of T_{SCM} and T_{CM} cells were very similar and could only be distinguished on the basis of eight DEGs (adjusted p-value < 0.01) (Fig 23A). One of these DEGs was *HNRNPLL*, which encodes heterogeneous nuclear ribonucleoprotein L-like, a master regulator of alternative splicing responsible for the expression of CD45RO[151]. In

contrast, CCR7⁺ PD-1⁺ TIGIT⁺ T_{PEX} cells were largely distinct, featuring lower expression levels of *SATB1*, which encodes a negative regulator of PD-1 expression[152], *MYC*, *DPP4*, which encodes CD26, *IL6ST*, *LEF1*, *IL6R*, and *NT5E* and higher expression levels of transcription factor (TF) genes recently associated with T

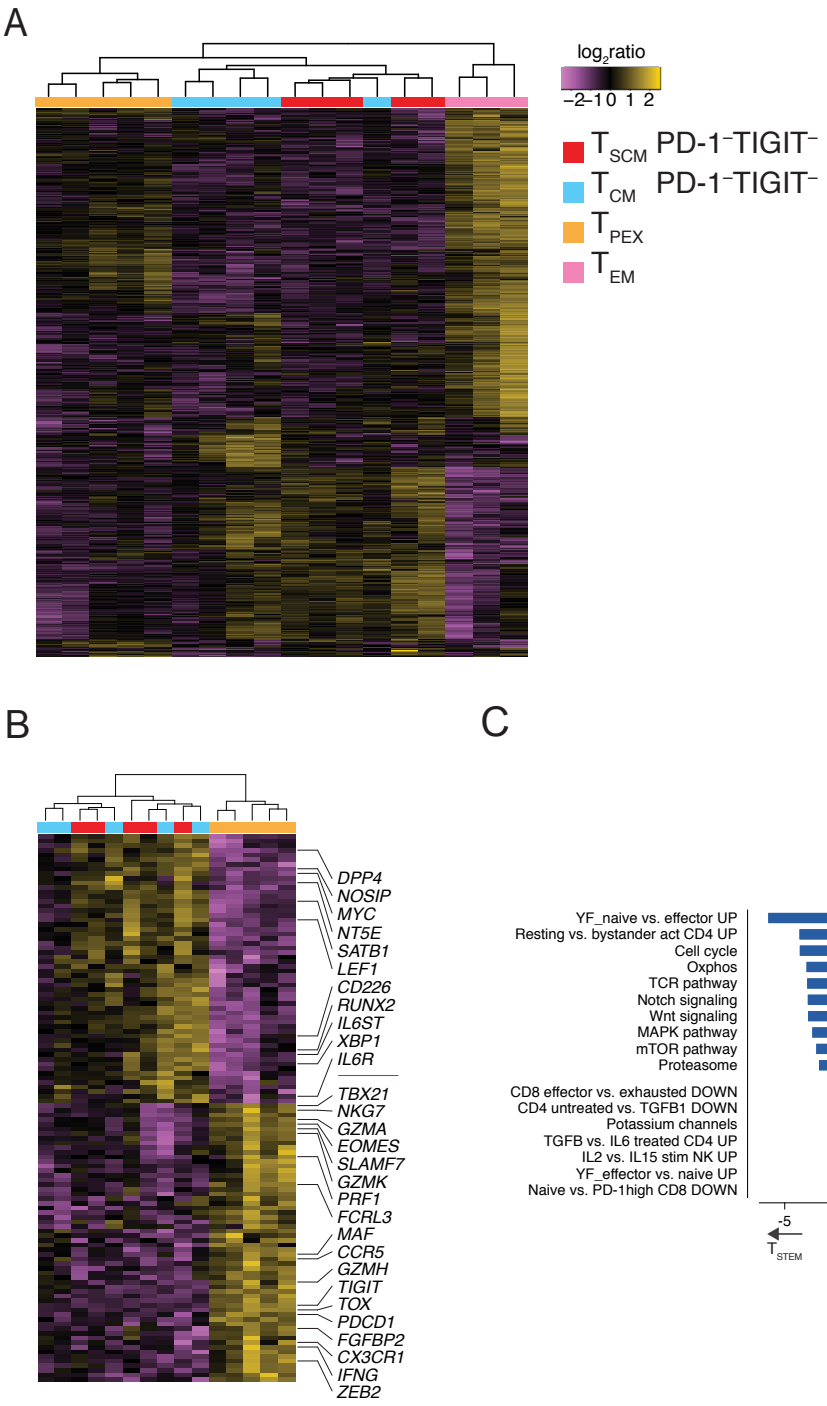


Figure 23 : T_{PEX} are characterized by a dysfunctional-like gene expression profile. Bulk RNAseq of $CD8^+$ memory subsets FACS-sorted from the peripheral blood of five HCs. Heat-map shows genes differentially expressed (Q value < 0.01). **B)** Heatmap showing differentially expressed genes (Q value< 0.01) by the indicated $CCR7^+$memory $CD8^+$ T cell subsets. Genes relevant for memory, effector or exhaustion differentiation are indicated. **C)** Histogram plot showing the normalized enrichment score (NES) of selected pathways obtained from GSEA in T_{STEM} versus T_{PEX} cells.

cell exhaustion, including *MAF*[153], *EOMES*[154] and *TOX*[99–102,155] and other genes associated with effector differentiation and cytolytic activity, including *ZEB2*, *GZMK*, *GZMA*, *TBX21*, *PRF1*, *IFNG*, and *NKG7* (Fig 23B). Moreover, GSEA further revealed that T_{STEM} cells were characterized by transcripts associated with the naive state, quiescence, oxidative phosphorylation, the Wnt[24] and Notch signaling pathways[156] and proteasome activity[157], whereas T_{PEX} cells were characterized by transcripts associated with the TGF-β signaling pathway[158], potassium regulation[159] and other mechanistic correlates of exhaustion, including the PD-1^{high} state. Transcripts associated with the cell cycle and the TCR and mTOR signaling pathways, collectively suggesting a predisposition to antigen-driven proliferation and effector differentiation, were also upregulated in T_{STEM} versus T_{PEX} cells (Fig 23C). In line with the transcriptional data, $CCR7^+$ PD-1^+ $TIGIT^+$ T_{PEX} cells stimulated with anti-CD3/CD28 and a combination of effector (IL-2 and IL-12) or homeostatic cytokines (IL-7 and IL-15) proliferated less vigorously than PD-1^- $TIGIT^-$ T_{SCM} and T_{CM} cells under identical conditions (Fig 24A). However, all three subsets proliferated similarly and remained phenotypically stable in response to IL-15, suggesting equivalent self-renewal capabilities. Accordingly, T_{SCM} and T_{CM} cells were better defined by the $CCR7^+$ PD-1^-

TIGIT⁻ phenotype, hereafter termed stem-like T (T_{STEM}), whereas early differentiated memory cells with exhaustion traits were characterized by the CCR7⁺ PD-1⁺ TIGIT⁺ T_{PEX} phenotype. **(**Fig 24B and C**)**

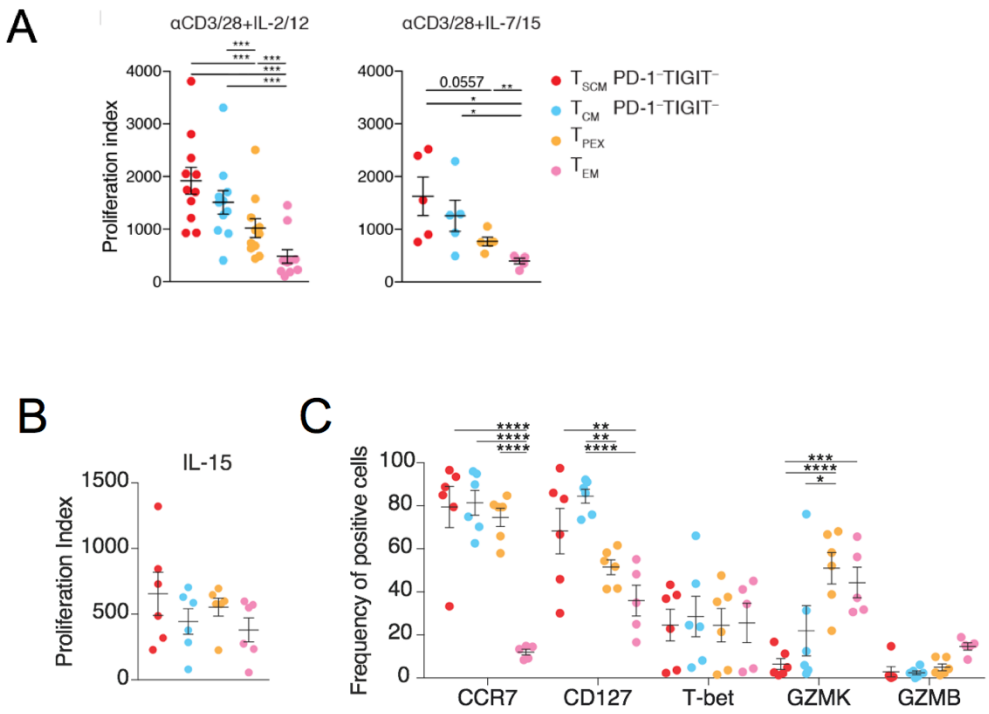


Figure 24 :Proliferative capacity of memory T cell subsets. A) Proliferation index of FACS-sorted memory subsets, obtained from HCs, stimulated with aCD3/CD28 (beads to cell ratio = 1:2) for TCR stimulation together with IL-2/IL-12 (10 ng/ml, n=11) or IL-7/IL-15 (10 ng/ml, n=5) for four days (n=10 and n=5, respectively. Proliferation index was calculated as Proliferation index = MFI non-proliferating fraction / MFI proliferating fraction × % cells with diluted CFSE. Data are shown as median. *P < 0.05; **P < 0.01; ***P < 0.001 (Paired t test and Wilcoxon test). **B)** Proliferation index of FACS-sorted memory subsets, obtained from HCs, stimulated with IL-15 (25ng/ml) for 10 days. Data are shown as in median. N=6. **C)** Histogram plot showing frequencies of markers expressed by the cells in **B**. Data are shown as median. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 (2way ANOVA).

4.2.2.2 T_{STEM} are functionally superior to T_{PEX} cells

Data obtained from transcriptional analysis suggest that T_{PEX} might feature a functional impairment in terms of proliferation, immune activation and effector function capacity compared to T_{STEM}. Thus we investigated functional properties of T_{PEX} and T_{STEM} both in vitro and in vivo.

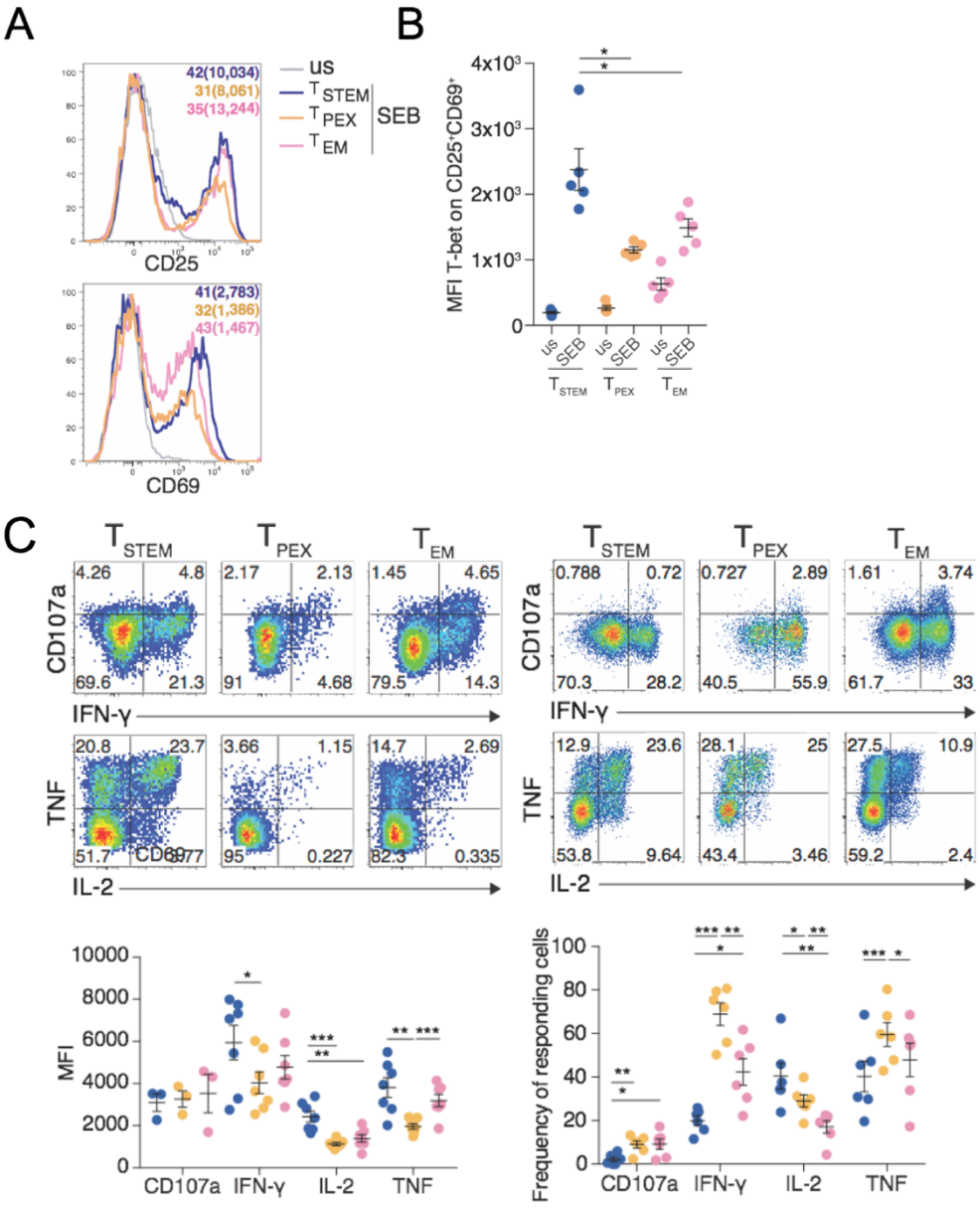


Figure 25 : Proliferation and effector function capacity of memory T cells subsets. A) Representative overlay of CD25 and CD69 expression by flow cytometry, from memory subsets stimulated with SEB (1ng/ml) for 24 hours. **B)** MFI of T-bet on activated cells (CD25+CD69+) for SEB stimulated subsets as in **A** or gated on total cells for unstimulated (Us) samples. Data are shown as median. N = 5. ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001 (Paired t test). **C)**Left, representative donor (top) and overall frequencies (bottom) of cytokine production of memory T cell subsets, stimulated for 12 hours with αCD3/CD28 in the presence of brefeldin A (1 µl/ml, Golgi Plug), monensin (0.67 µl/ml, Golgi Stop). Data are shown as median. N=7. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 (Paired t test). Right, representative donor (top) and overall frequencies (bottom) of cytokine production of memory T cell subsets FACS-sorted as in **A** and **B**, stimulated for 3 hours with PMA/Ionomycin in the presence of brefeldin A (1 µl/ml, Golgi Plug), monensin (0.67 µl/ml, Golgi Stop). Data are shown as median. N=6. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 (Paired t test).

In response to TCR-dependent stimulation with Staphylococcal enterotoxin B (SEB), FACS-purified T_{STEM} cells upregulated CD25 and CD69 to a greater extent than T_{PEX} cells (Fig 25A), and activated $CD25^+$ $CD69^+$ T_{STEM} cells expressed higher levels of T-bet than activated $CD25^+$ $CD69^+$ T_{PEX} cells (Fig 25B). Likewise, T_{STEM} cells produced cytokines at higher levels on a per cell basis (IFN-γ, IL-2, and TNF) than T_{PEX} cells in response to stimulation with anti-CD3/CD28 (Fig 25C, left panels). No clear differences were observed between T_{STEM} and T_{PEX} cells with respect to degranulation, measured via the surface mobilization of CD107a (Fig 25C, left panels). In response to TCR-independent stimulation with phorbol myristate acetate (PMA) and ionomycin, however, T_{PEX} cells produced IFN-γ and TNF and mobilized CD107a at much higher frequencies than T_{STEM} cells, the functional superiority of which was therefore limited to conditions that mimicked antigen recognition events (Fig 25C, right panels).

To achieve further information regarding differentiation status of T_{STEM} and T_{PEX} we investigated telomere length, according to Single Telomere Length Analysis (STELA) (Fig 26). Measurement of telomere length displayed high similarity

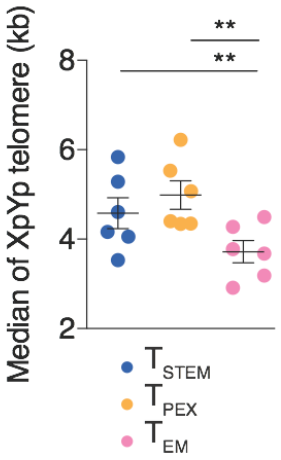


Figure 26 : Telomere length of memory T cell subsets. 6,000 T_{STEMCM}, T_{PEX} and T_{EM} cells were sorted from total PBMCs of six healthy donors and STELA assay was performed. Median telomere length of each subset is shown in figure. Data are shown as median. **P < 0.01 (RM one-way ANOVA).

between subsets, indicating that both T_{PEX} and T_{STEM} underwent similar number of cell divisions. We finally sought to test the long-term repopulation capacity of these subsets by performing serial adoptive cell transfers (ACTs) in NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) humanized mice in vivo (Fig 27A). Thus we FACS-sorted T_{PEX}, T_{STEM} and T_{EM} from healthy controls and adoptively transferred in recipient mice alongside CD8-depleted autologous PBMCs as support. Purified T_{EM} failed to efficiently repopulate these mice following the first ACT (not shown), in line with previous data [25]. While both T_{STEM} and T_{PEX} expanded to similar numbers in primary hosts (not shown), T_{STEM} proliferated faster in the

blood (Fig 27B) and preferentially repopulated the spleen compared to T_{PEX} at 4 weeks after retransfer in secondary hosts (Fig 27C).

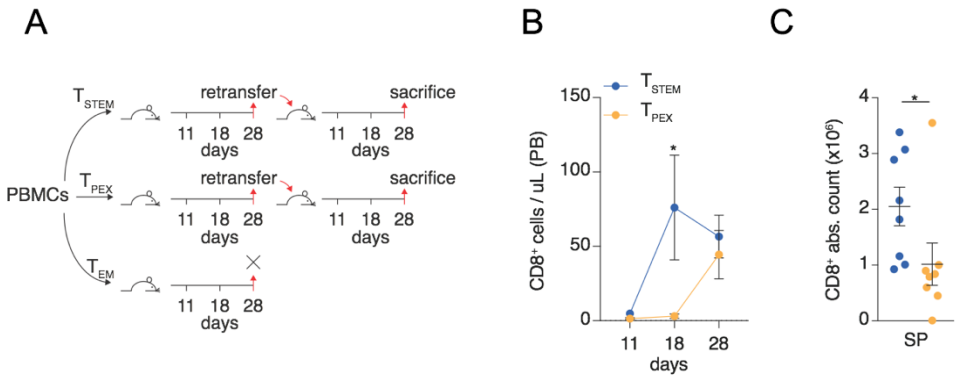


Figure 27 :T_{PEX} display lower expansion capacity upon serial transfers in vivo A) Experimental layout of serial transplantation experiments. FACS-purified T cell subsets were transferred retroorbitally in primary NSG recipients. 28 days post transfer, human T cells were isolated from spleen and lungs, the CD4:CD8 ratio normalized and T cells were subsequently transferred in secondary recipients. **B)** Absolute number of $CD8^+$ T cells/µL in the peripheral blood (PB) at day 11, 18 and 28 after transfer of T_{STEM} and T_{PEX} cells in secondary recipients. **C)** Total $CD8^+$ T cells recovered in the spleen (SP) of secondary recipient mice treated as in H (day 28 post transfer). Data in I and J indicate n=4 mice/group; two pooled independent experiments). *P < 0.05 (2-way ANOVA and Mann-Whitney test). In both graphs, data are shown as mean ± SEM.

4.2.3 T_{PEX} are committed to a terminally dysfunctional state

Epigenetic regulation plays a key role in T cell fate decisions[93]. We therefore employed the Assay for Transposase-Accessible Chromatin using sequencing (ATACseq) to compare the open chromatin landscapes of T_{STEM} and T_{PEX} cells in terms of differentially accessible regions (DARs).

Naive and T_{EM} cells were analyzed in parallel as lineage controls. PCA in Fig 28A revealed

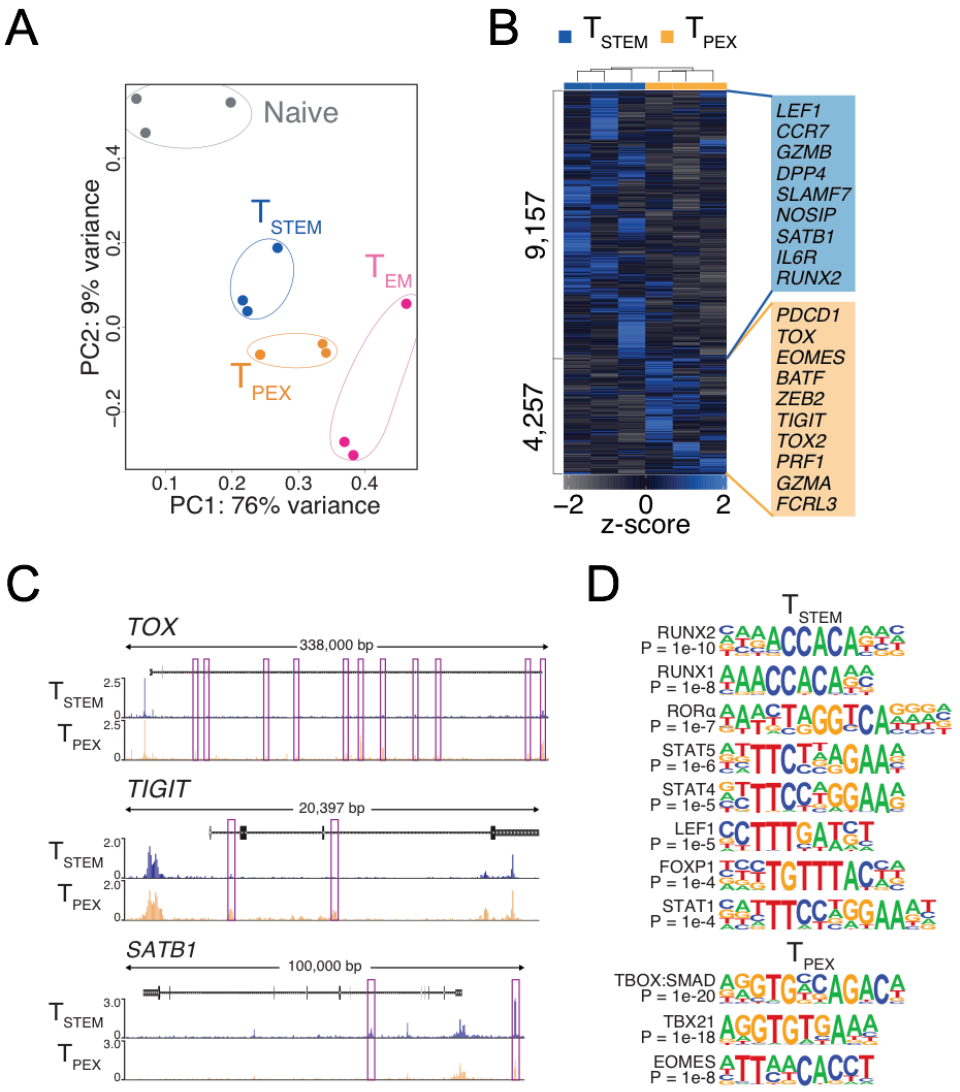


Figure 28 : Chromatin accessibility analysis of memory T cell subsets. T_N, T_{STEM}, T_{PEX} and T_{EM} cells were FACS-sorted from 3 HC's peripheral blood and processed for ATACseq library generation and sequencing. **A)** Principal component (PC) analysis showing the distribution of the samples along the PCs. The numbers indicate the DARs between each subset. **B)** Heatmap showing ATAC-seq DARs. Genes relevant for memory, effector and exhaustion differentiation that show DARs are depicted in the yellow (open in T_{PEX}) or blue (open in T_{STEM}) box. **C)** representative genomic regions showing the chromatin status for TOX, TIGIT and SATB1 in the two subsets in **B**. The peaks that are differentially open are highlighted in red. **D)** Transcription factor binding motifs enriched in DARs shown in B. The P value for each enrichment is indicated.

that T_{STEM} and T_{PEX} cells were globally similar, although T_{STEM} cells mapped toward the naive subset, whereas T_{PEX} cells mapped toward the T_{EM} subset (Fig 28A). However, we also identified a total of 13,414 DARs between T_{STEM} and T_{PEX} cells (Fig 28B). Genes associated with T cell dysfunction (*e.g.*, *TOX*, *TOX2*, *TIGIT*, *PDCD1*, *NFATC2*, and *MAF*), terminal differentiation (*e.g.*, *ZEB2* and *BATF*), and other immune-related processes previously identified at the mRNA level (*e.g.*, *EOMES* and *GZMA*) were more accessible in T_{PEX} versus T_{STEM} cells (Fig. 28B and C). In contrast, genes associated with T cell memory (*e.g.*, *LEF1*, *SELL*, *CCR7*, *BACH2*, and *SATB1*) and effector functions [*e.g.*, *GZMB*, *RORA*[160] (Fig 28B and C). Computational analysis by HOMER in these DARs identified TF binding motifs (TFBMs) preferentially enriched in T_{STEM} vs. T_{PEX}, and related to TFs generally expressed by less differentiated T cells (RUNX2, RUNX1, LEF1 and FOXP1), involved in effector differentiation (RORA) or mediating cytokine signaling (STAT5, STAT4 and STAT1). Instead, those motifs preferentially enriched in T_{PEX} were related to TFs involved in effector differentiation but also in the specification of exhausted cells, such as TBX21 (encoding T-bet) and EOMES. The combined TBOX:SMAD motif was also identified in these cells (Fig 28D). The chromatin accessibility data suggested that T_{PEX} cells were predisposed to the generation of dysfunctional progeny and susceptible to the inhibitory effects of TGF-β signaling via SMADs. Accordingly, T_{PEX} cells proliferated to a lesser extent and produced less GZMB than T_{STEM} cells in response to stimulation with anti-CD3/CD28 and IL-15 (Fig. 29), recapitulating our previous data on the functional defects of T_{PEX} cells also in this stimulation condition. The addition of TGF-β further inhibited these responses, especially the production of GZMB, in parallel cultures of T_{PEX} cells,

whereas minimal effects were observed in parallel cultures of T_{STEM} cells (Fig. 29).

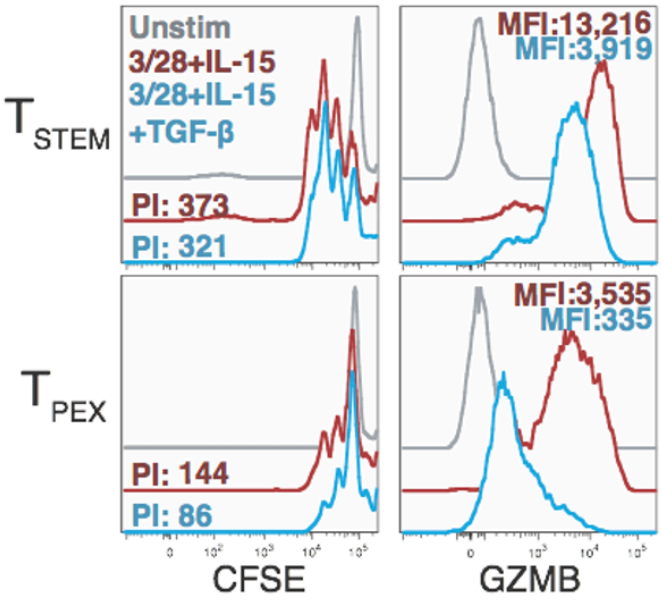


Figure 29: T_{PEX} display dysfunctional traits after TCR stimulation. FACS-sorted memory subsets stimulated in vitro with αCD3/CD28 (beads to cell ratio = 1:2) together with IL15 (10 ng/ml) alone or in combination with TGFβ (10 ng/ml). The histograms show the overlay from unstimulated (Unstim), aCD3/CD28 + IL15 and αCD3/CD28 + IL15 + TGF-B stimulated samples from one representative donor. The proliferation index was calculated as in Fig **21A.**

Eventually, from the same dataset we also investigated multipotency of T_{STEM}, originally FACS sorted as PD-1$^-$TIGIT$^-$, and T_{PEX}, originally isolated as PD-1$^+$TIGIT$^+$. Notably, T_{PEX} cells generated a progeny that almost exclusively kept the same PD-1$^+$TIGIT$^+$ phenotype, while T_{STEM} resulted capable to generate all the phenotypical combinations of PD-1 and TIGIT, thus displaying increased plasticity compared to T_{PEX}, which on the

opposite are strictly committed to generate a dysfunctional progeny in response to TCR stimulation (Fig 30).

We then used RNAseq to profile the transcriptomes of T_{STEM} and T_{PEX} cells after stimulation with anti-CD3/CD28 in the presence of IL-2 and IL-12 (Fig 31A). Activated T_{STEM} preferentially expressed the memory-related

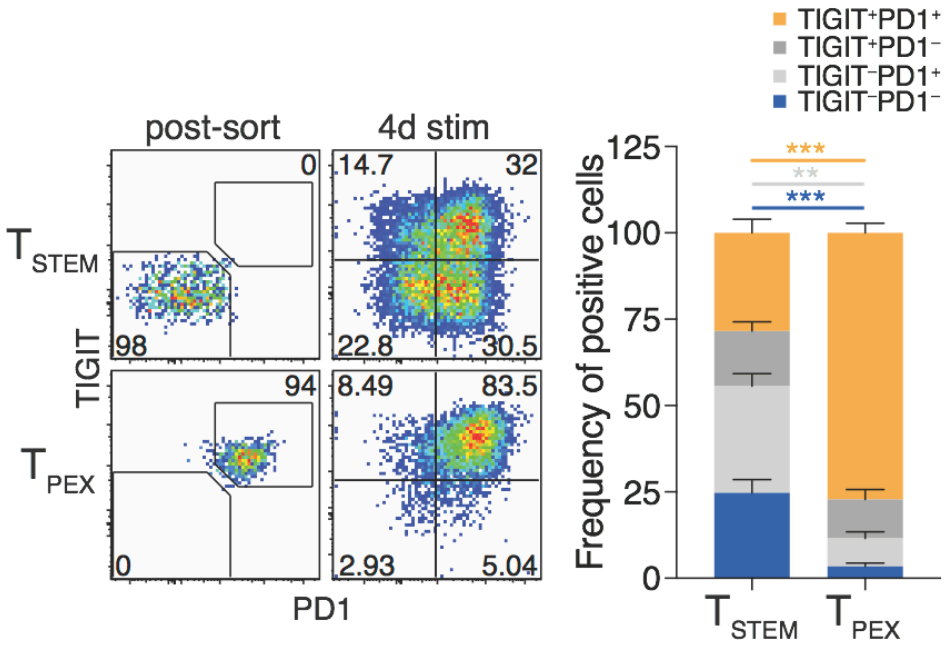


Figure 30 : Differentiation potential of T_{STEM} and T_{PEX}. Dot plots showing the phenotype of T_{STEM} and T_{PEX} after sort and after four days of stimulation from one representative donor (left). Bar plot show pooled results for the T_{STEM} and T_{PEX} progeny cell phenotype from all donors. Data are shown as median. N = 5. **P < 0.01; ***P < 0.001 (Unpaired t test and Mann-Whitney test).

genes *BACH2*, *ID3*, *IL2* and *SATB1* alongside the effector-related genes *IRF8*, *RORC*, *GNLY*, *XBP1*, *IL26* and *IL23R*, whereas activated T_{PEX} overexpressed genes related to dysfunctional cells such as *TOX*, *PDCD1*, *TIGIT*, *MAF* and *CXCL13,* the inhibitor of IL-2 production *IKZF3*, the TGFβ -induced TF *SMAD3*, the cytotoxicity-related molecules *GZMK*, *GZMH* and *GZMA*, as well as additional immune-related molecules, including multiple chemokines.

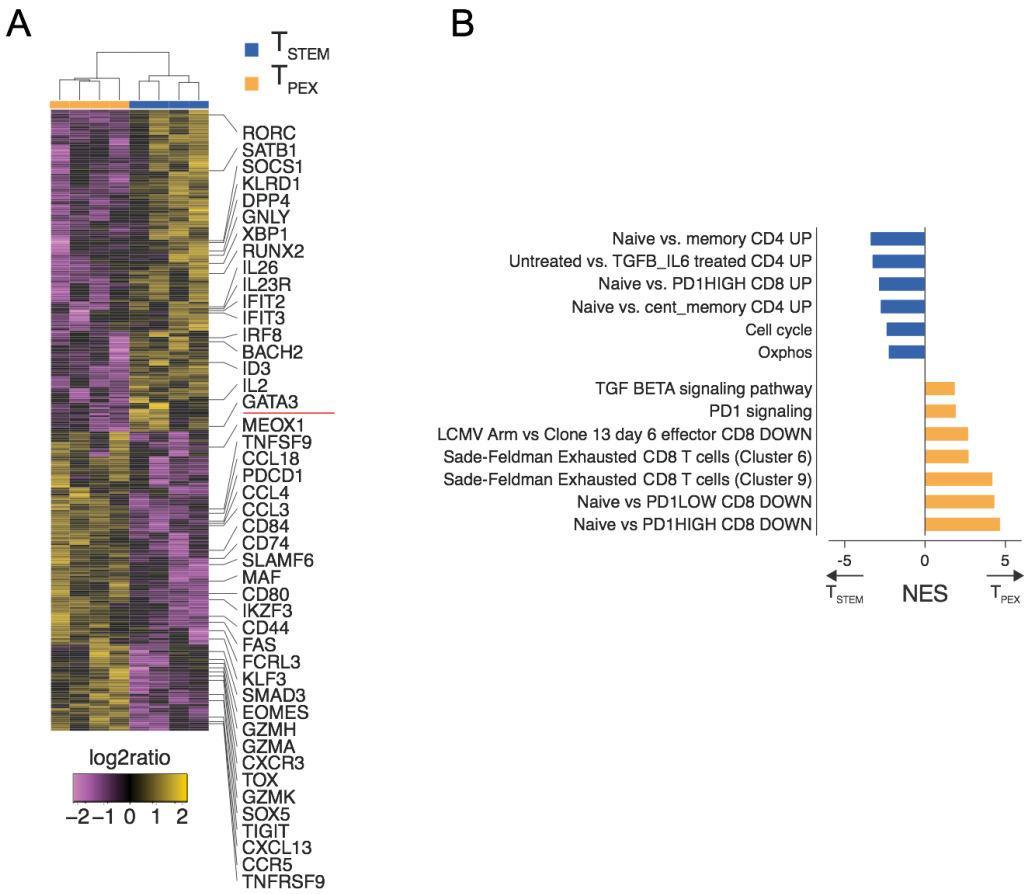


Figure 31 : Transcriptome analysis of activated T_{STEM} and T_{PEX} . A) Heatmap showing selected differentially expressed genes (Q value<0.05) by FACS-sorted T_{STEM} and T_{PEX} stimulated as in Fig 28, **B)** Histogram plot showing the normalized enrichment score (NES) of selected pathways obtained from GSEA in activated T_{STEM} versus T_{PEX} cells. Selected pathways are shown (Q value <0.05).

Notably, some of these genes were previously found differentially expressed ex vivo (Fig. 23B). GSEA further revealed the functional vs. dysfunctional nature of T_{STEM} and T_{PEX}, respectively, according to which T_{STEM} were preferentially enriched in gene sets related to less differentiated, non-exhausted or proliferating T cells, while T_{PEX} in gene sets related to TGFβ and PD-1 signalling, PD-1^{high} cells or exhausted T cell subsets identified in melanoma tumors by scRNA-seq [89] (Fig 31B). PCA of ATAC-seq data from paired samples showed that stimulation induced substantial changes in the

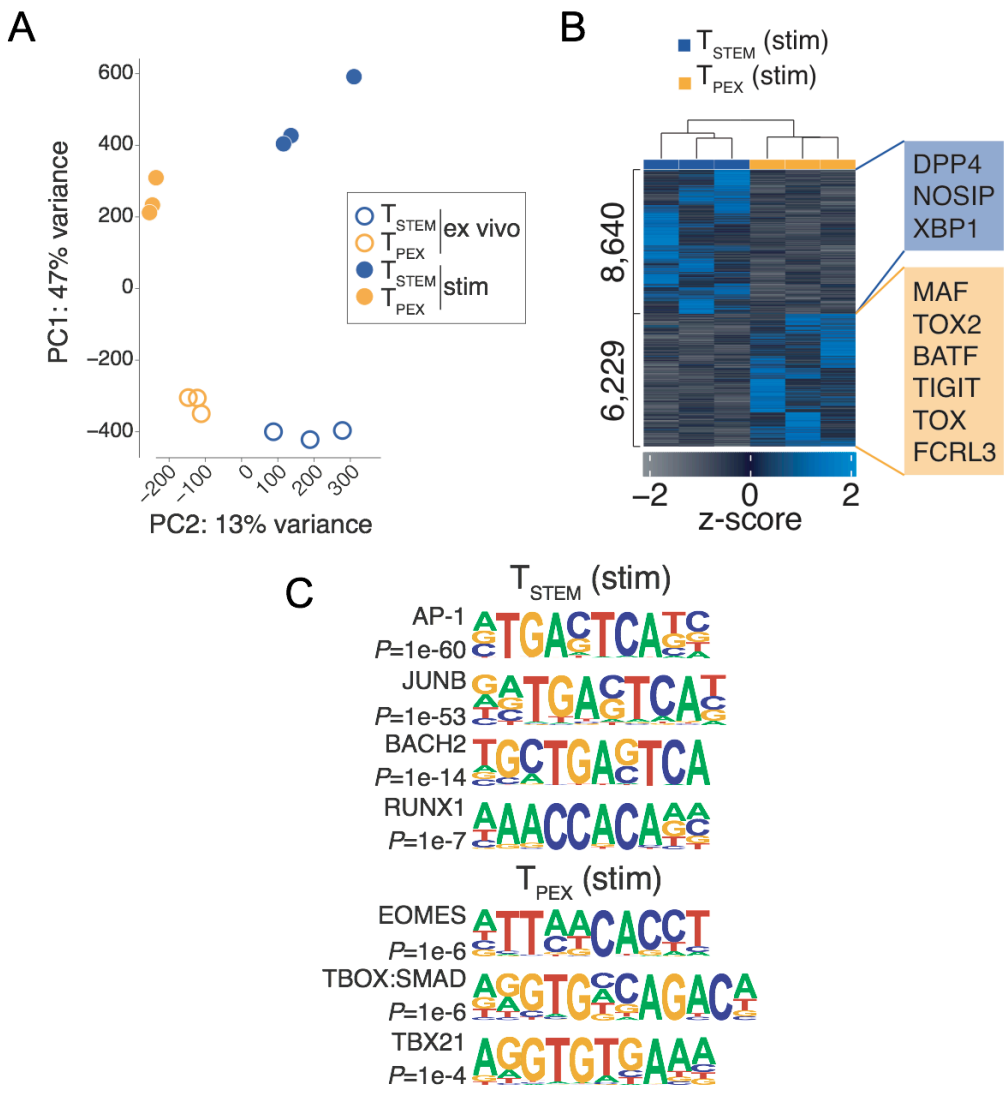


Figure 32 : Epigenetic landscape of activated T_{STEM} and T_{PEX} . A) PCA as in Fig 25A of FACS sorted T_{STEM} and T_{PEX} stimulated as in Fig 26 or left unstimulated. **B)** Heatmap showing DARs between activated T_{STEM} and T_{PEX} cells. Genes relevant for memory, effector and exhaustion differentiation that show DARs are depicted in the yellow (open in T_{PEX}) or blue (open in T_{STEM}) box. **C)** Transcription factor binding motifs enriched in DARs between activated T_{STEM} and T_{PEX} cells. The P value for each enrichment is indicated.

chromatin accessibility landscape of both T_{STEM} and T_{PEX} compared to controls ex vivo (Fig 32A). However, those major differences previously identified ex vivo and visible along PC2 were stably maintained both at the level of specific genes (Fig 32B) and of TFBM enrichment (Fig 32C),

thereby indicating that the two subsets follow distinct paths of differentiation that are epigenetically imprinted.

Finally, we reasoned that the differences in effector, memory and exhaustion traits illustrated above might impact the functionality of T cells in vivo. In this regard, we used a stringent ACT protocol where cell products derived from naturally-occurring T_{STEM} and T_{PEX} and redirected with a chimeric antigen receptor (CAR) targeting CD19 are transferred in tumor-bearing mice in the absence of human autologous $CD4^+$ T cells or exogenous cytokines. We noticed that T_{STEM} displayed enhanced control of leukemic burden compared to T_{PEX} at multiple time points following ACT in NSG mice previously injected with the acute lymphoblastic leukemia NALM6 cell line (Fig 33). Overall, our results indicate that T_{PEX} are "hard-wired" to their dysfunctional signature after immune activation and effector differentiation, while T_{STEM} are relatively resistant to exhaustion, thereby resulting in enhanced anti-tumor potency *in vivo*.

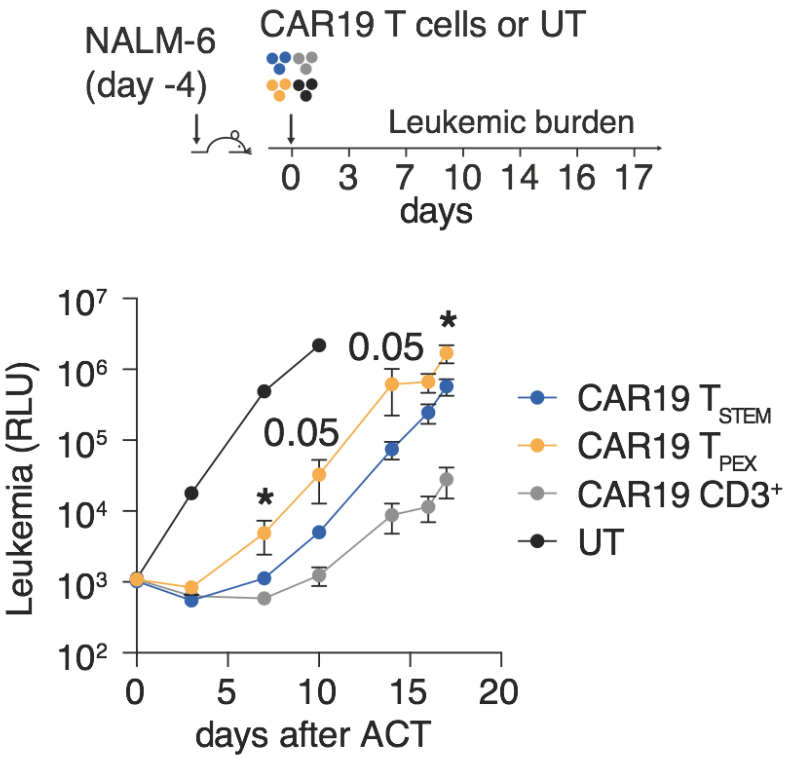


Figure 33 : T_{STEM} display increased capacity to control tumor burden in vivo. Top: schematic layout of the adoptive transfer experiment. Bottom: relative luminescence units (RLU) measured overtime and indicating the growth of the NALM6 cells in NSG mice adoptively transferred with T cell subsets previously redirected with anti-CD19 chimeric antigen receptor (CAR19). Follow-up was stopped when >75% of mice from one of the treated group reached an RLU value ≥10^6. Data are shown as mean ± SEM. UT: unstransduced $CD3^+$ T cells. *P < 0.05 (Unpaired t test and Mann-Whitney test).

4.2.4 T_{PEX} cells are abundant in persistent infections and clonally distinct from T_{STEM} cells

Dysfunctional, exhausted $CD8^+$ T cells develop in response to persistent TCR stimulation[161] thus we reasoned that T_{PEX} cells could be more abundant in T cells specific for chronic versus acute infections. Recent works demonstrated that a single round of vaccination is known to induce a population of long-lived T_{SCM}-phenotype cells in response to yellow fever virus (YFV)[162,163]. As we have shown that T_{SCM}-phenotype cells are heterogeneous, we tested whether the vaccine-induced YFV-specific $CD8^+$ T cells preferentially belong to the functional T_{STEM} or dysfunctional T_{PEX} phenotype. In order to do this, we took advantage of publicly available gene expression data from vaccinated individuals. In this regard, we found that the YFV-specific T_{SCM}-cells persisting years after vaccination shared a transcriptional identity with T_{STEM} but not with T_{PEX} cells (Fig 34A). In addition, we used MHC class I multimers to investigate

antigen-specific CD8⁺ T cells in response to different persistent (CMV and EBV) or non-persistent (Flu and Rotavirus)-related epitopes in 5 individuals (2 of whom HIV⁺) by mass cytometry (CyTOF). As depicted in Fig 34B, We found that the CCR7⁺ fraction of CMV and EBV-specific CD8⁺ T cells had increased levels of the T_{PEX} signature markers TIGIT, PD-1, GZMK, GZMA and CCR5 compared to that of the Flu-specific CD8⁺ T cells. This was more evident for EBV than CMV-related epitopes. We

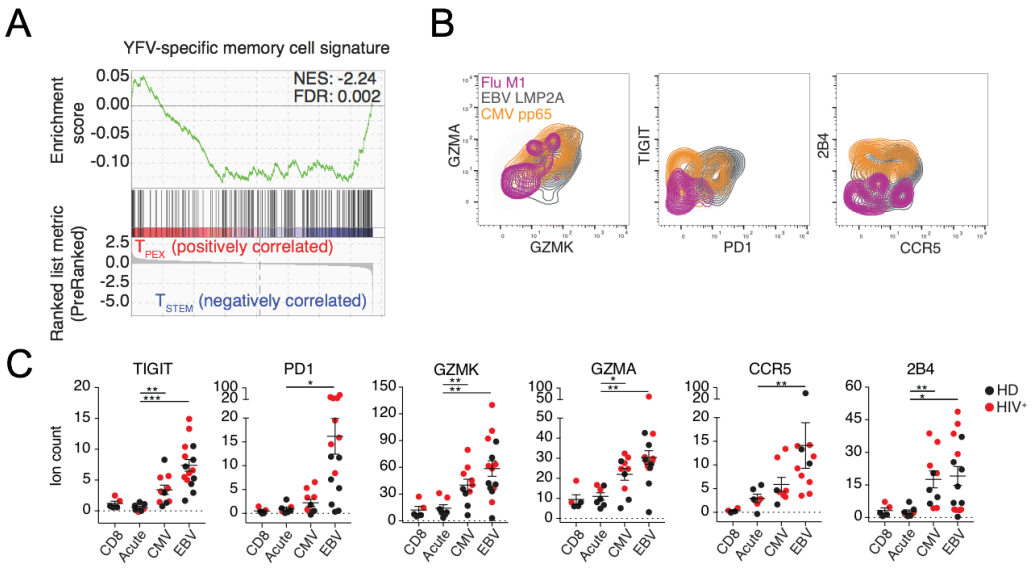


Figure 34 : T_{PEX} are more abundant in chronic Ag-specific memory pool. **A)** GSEA plot of YFV-specific CD8 T cell signature, derived from public gene expression data, in T_{STEM} versus T_{PEX} cells. **B)** CyTOF plots showing overlay of total memory cells with T_{STEM}, T_{PEX} and CCR7⁺ virus-specific fractions for Flu M1, EBV LMP2A and CMV pp65 epitopes. **C)** Ion count quantification by CyTOF of the markers analyzed in B. Flu (n=6) and Rotavirus (n=1) epitopes were labelled as “Acute” for simplicity. Each category (“Acute”, “CMV” and “EBV”) includes Epitope-specific responses from a single donor (for a total of 5 donors) directed to the same virus were labelled as “CMV” or “EBV”. Total CD8 were also reported as a reference. *P < 0.05; **P < 0.01; ***P < 0.001 (Unpaired t test and Mann-Whitney test).

additionally found overexpression of the 2B4 receptor by cells specific for persistent viruses, suggesting that 2B4 can serve as an additional marker for their identification. No relevant differences were found between healthy

and HIV$^+$ donors. (Fig 34B and C) The abundance of T_{PEX} cells was similar when measuring CMV-specific CD8$^+$ T cells with high avidity for the TCR by the use of a point-mutated MHC class I tetramer[164] thereby corroborating the recent idea that persistent stimulation rather than TCR signal strength generates dysfunctional CD8$^+$ T cells (Fig 35). Overall, these data indicate that persistent antigen stimulation, even in healthy

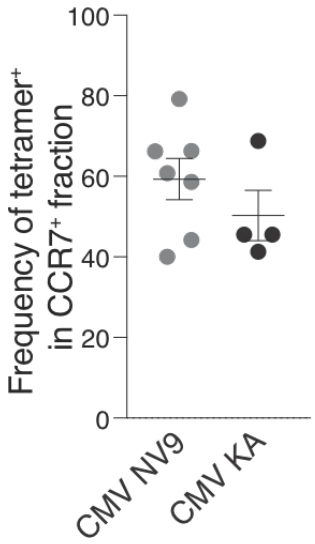


Figure 35 : T_{PEX} generation is not shaped by TCR avidity. Frequency of Ag-specific T_{PEX} for HLA-A*0201/restricted CMV pp65$_{495-503}$ NLVPMVATV (NV9) and CMV KA$_{227/8}$ NLVPMVATV (KA) tetramers. n=7 (pp65) and 4 (KA).

individuals, is preferentially associated with the development of a phenotype endowed with diminished functionality. We next performed deep sequencing of the TCR (TCRseq) in FACS-purified cells from 6 different individuals to corroborate the evidence that T cell responses contained in T_{STEM} and T_{PEX} memory subsets are directed towards

different epitopes. T_{EM} cells were also sorted in parallel to test diversity and overlap of the TCR repertoire as a control. In line with their less-

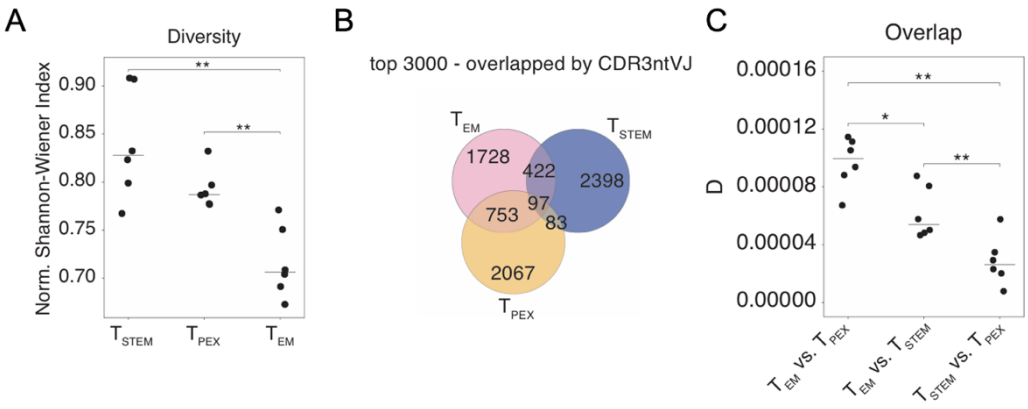


Figure 36: TCR repertoire analysis display little overlap between T_{STEM} and T_{PEX} A) Normalized Shannon-Wiener index indicating the clonal diversity of the TCR repertoire for each population as determined by TCRβ chain deep sequencing. **B)** Summary of the pairwise comparison between subsets of the data in A (n=6). In both A and B, the horizontal bar indicates the median. *P < 0.05, **P < 0.01; Diversity (D) metrics. **C)** Venn diagram shows overlap of top 3000 TCR clonotypes shared between T_{STEM}, T_{PEX}, and T_{EM} subsets from a representative donor.

differentiated memory identity, T_{PEX} appeared as polyclonal as T_{STEM} memory cells when compared to the less diverse, oligoclonal T_{EM} cells (Fig 36A) (analysis related to the most abundant 3,000 clones). Moreover, while both shared clones with T_{EM} cells, T_{STEM} and T_{PEX} showed little overlap between each other (Fig 36B and C), suggesting a divergent differentiation program capable to give rise to more differentiated effector T cell progeny.

5 Discussion

The $CD8^+$ T_N cell compartment has been shown during last years to be composed by distinct subsets, including cells that display clear memory traits despite their naïve-like phenotype, such as T_{SCM}[25,30] and $CXCR3^{bright}$ T_{MNP} cells[141]. With this project, we demonstrated that truly naïve human $CD8^+$ T cells, as defined by a plethora of assays, possess distinct features. We have identified two populations of $CD8^+$ T_N cells on the basis of the expression of CXCR3, defined as T_NR3^- and T_NR3^+.

T_NR3^+ displayed increased cytokine production capability, both after mitogen and cognate antigen stimulation in vitro, augmented effector differentiation potential after foreign antigen encounter, both in vivo and in vitro, and overexpressed a different set of genes compared to T_NR3^-. These features correlated with physico-chemical properties of TRB CDR3, thus suggesting that T_NR3^+ cells bear TCRs with equivalent diversity compared to T_NR3^-, but increased strength of interaction with p:MHC complexes. Furthermore, human T_NR3^+ cells display a shared transcriptional identity with their murine counterparts, which are comprised in the $CD5^{hi}$ fraction of $CD44^{lo}$ $CD62L^{hi}$ T_N cells. In the mouse, the potential to respond to foreign antigen correlates with TCR sensitivity for self-antigens[72], and such sensitivity in turn correlates with surface CD5 expression[70–73]. Thus, FACS-sorted $CD5^{hi}$ T_N cells, that included T_NR3^+ cells[72], displayed a gene expression profile featuring increased levels of transcripts related to cell cycle and effector differentiation, such as *Tbx21*, *Eomes* and *Cxcr3*, compared to $CD5^{lo}$ cells[72]. In fact, according to gene expression data, a small subset of murine $CD5^{hi}$ also expresses CXCR3[72] (Fig 16A). In our hands, these murine T_NR3^+ cells revealed increased expression of genes related to effector differentiation compared to both

$CD5^{lo}$ and $CD5^{hi}$ cells, also displaying a transcriptional identity shared with human T_NR3^+.

The human TCR repertoire is highly polyclonal, thus to further investigate the potency of immune response we took advantage of the mouse T_N pool, where well known affinity for self-derived Ags can be exploited. In the context of transgenic TCR, murine T_NR3^+ cells displayed slightly, but significantly increased capability to generate $KLRG^{hi}$ IL-$7R^{lo}$ SLECs following acute LCMV infection. Interestingly, even in the presence of normalized TCR, T_NR3^+ cells displayed to be poised to effector differentiation, thus suggesting an intrinsic pre-commitment of T_N subsets. Collectively, these murine data indicate that murine T_N cell subsets that are equivalent to those from humans are differentially committed to generate effector T cell progenies following activation, thereby corroborating the evidence that effector differentiation potential is shaped at the level of the preimmune repertoire.

CD5 has been recently used as an indicator of human $CD8^+$ T cell affinity for self[74]. However, we have shown that CD5 expression is not capable to segregate subsets of T_NR3^- and T_NR3^+ (Fig 16A), thus questioning its role in defining T_N heterogeneity in humans. Taken together, these observations and the differential bias for effector differentiation correlating with CXCR3 expression suggest that CXCR3 is a more efficient marker to identify T_N cells with increased effector capacity.

Recently, work from Alanio and colleagues[165] demonstrated that memory-phenotype $CD8^+$ T cells display a mechanism of retention in the thymus and the skin which is driven by CXCR3/CXCL10 axis in unimmunized mice. According to our data, T_NR3^+ are not present in mucosal tissues (i.e skin gut and lung, Fig 7A) and do not preferentially home to specific

organs in the periphery. On the other hand, they are relatively abundant in PB and LNs. Although in the mouse there is previous evidence of T_N-phenotype or pre-activated T_N cells entering non-lymphoid organs (i.e. lung, gut) [166,167], our data are in line with a recent report by Thome et al, who have shown the virtual absence of T_N cells in human mucosal tissues[133]. Work from Annunziato and colleagues[168] demonstrated that CCR7 and CXCR3 expression are mutually exclusive within human thymocytes. Additionally, previous work from our collaborator revealed that CXCR3 signaling is absent in $CD8^+$ T_N and T_{MEM} at the steady-state but increased with TCR triggering[169]. Furthermore, also chemokines ligand of CXCR3 are present at low levels at the steady state, while increase with inflammation[170,171].These observations, together with the absence of the tissue-residency markers CD69 and CD103, suggests a constant recirculation between PB and LNs of T_N subsets and argue against a role of CXCR3 in T_N cell physiology and homeostasis in the absence of cognate Ag stimulation. According to our data, T_NR3^+ might have a migratory advantage over T_NR3^- only after T cell activation, in line with what has been reported for murine T_{CM}, which are capable of rapid localization to peripheral areas of LNs after Ag exposure thanks to their preferential expression of CXCR3[171].

T_NR3^+ cells display only a modest 2-fold dilution in TREC content compared to T_NR3^-, which represent an average of one cell division. In our opinion, this small entity dilution is not sufficient to allow the identification of $CD8^+$ counterpart of $CD4^+CD31^+$ Recent Thymic Emigrants (RTE), which instead display a TREC content on average 8 times higher than $CD4^+CD31^-$[172]. However, TREC dilution of T_NR3^+ cells suggests a higher turnover due to homeostatic proliferation, in line with what observed by

Cho and colleagues, which demonstrated that murine $CD5^{hi}$ responded better to lymphopenia-induced proliferation than $CD5^{lo}$ T_N cells[73]. According to their naive status, CXCR3 upregulation in T_NR3^+ cells is not driven by antigen recognition. Instead, increased homeostatic proliferation of T_NR3^+ cells might play a role in this regard. In particular, increased tonic signaling due to distinct TCRs with intrinsic affinity for self-Ags may cause a higher rate homeostatic proliferation which can in turn drive CXCR3 upregulation. This speculation is in line with recent findings from Kato and colleagues, which demonstrated the conversion of T_NR3^- cells in T_NR3^+ at slow cell division rate in lymphopenic mice[173]. Interestingly, recent work from Smith et al supports this hypothesis. Indeed, they demonstrated that the murine T_N pool is organized in developmental layers in response to infection, with a fetal layer of cells, which displayed some degree of overlap with T_NR3^+, prone to effector differentiation while cells composing the adult layer tended to generate a long-lived memory progeny[174]. Collectively, with this project we demonstrated that the human and murine T_N compartments are composed by at least two subsets which can be defined on the basis of CXCR3 expression. T_NR3^+ are characterized by enhanced capability to differentiate in functional effector cells, thus providing a selective advantage during an immune response. It is still unclear why a subset prone to generate a less efficient progeny, i.e., the T_NR3^- cells is evolutionary maintained, but it is tempting to speculate that their persistence may avoid the generation of “holes” in the TCR repertoire that pathogens could potentially exploit. Thus, T_NR3^- cells could represent an efficient strategy to guarantee a polyclonal immune response, albeit at lower magnitude[175].

Overall, with this project we brought attention to the quality of preimmune repertoire as an additional layer of functional and transcriptional regulation of effector T cell responses in humans.

In the second part of the project, we have better defined long-lived memory T cell differentiation, where stem-like $CD8^+$ memory T cells with strong proliferative capacity are identified as $CCR7^+$ $PD\text{-}1^-TIGIT^-$ independently of their T_{SCM} or T_{CM} phenotype. Prior to our study, the discovery of T_{SCM} defined this subset as a progenitor for the memory compartment equipped with enhanced proliferative capacity and anti-tumor properties[25]. Conversely, T_{CM} present higher cytotoxic properties compared to T_{SCM} but lower multipotency and persistence. As a consequence, T_{SCM} were proposed as the most promising cell product for cancer immunotherapy[176]. In light of our data, the differences previously ascribed to T_{SCM} and T_{CM} can be largely explained to a differential contamination of the transcriptionally and functionally distinct T_{PEX} cells within these subsets. Our results suggest that the generation of a cell product with a defined $PD\text{-}1^-TIGIT^-$ phenotype from T_N precursors, independent from previous T_{SCM} and T_{CM} definition, will enhance the beneficial properties of adoptive transfer of CAR-engineered T cells in cancer treatment. Indeed, this will provide a population of T cells ($CCR7^+GZMK^-PD\text{-}1^-TIGIT^-$, T_{STEM}) with improved self-renewal, proliferative and cytotoxic potential. Additionally, the identification of a larger pool of progenitors with enhanced stem-like properties, the T_{STEM}, compared to T_{SCM} cells, could overcome the limitation of small numbers to be redirected with CARs ex vivo.

Moreover, we characterized a subset of T cells that we defined T_{PEX}, simultaneously featuring long-lived memory and dysfunctional traits at the

gene expression, epigenetic and functional level, resembling stem-like progenitors of exhausted T cells recently identified in the context of chronic viral infections and cancer. Earlier studies defined a dichotomy within the T_{EX} compartment, according to which terminally differentiated $TCF1^-$ and progenitor $TCF1^+$ subsets can be identified[82,83,88,177]. On the basis of these data, it has been proposed that exhaustion differentiation mirrors memory T cell differentiation despite their developmental relationships remain unclear[75,178]. Our data on clonal divergence of T_{STEM} and T_{PEX} cells strongly supports the concept that layers of memory T cells exist in humans with parallel programs of differentiation. The recent identification of TOX as a specific driver of exhaustion via epigenetic modifications[99–102,155] further confirms that memory and exhaustion differentiation are governed by distinct mechanisms. In humans, T_{PEX} express increased levels of *TOX* mRNA and show preferential accessibility of the *TOX* gene compared to T_{STEM}. On a functional level, this reflects on lower proliferation and production of cytokines such as IFN-γ, IL-2 and TNF-α compared to T_{STEM}, although we did not formally address whether this is due to TOX or to expression of additional factors previously linked to exhaustion. Moreover, the sensitivity to the suppressive cytokine TGFβ corroborated the predisposition to dysfunction by T_{PEX} compared to T_{STEM} cells. Interestingly, T_{PEX} express markers that are important for long-lived memory T cell homeostasis (e.g., CCR7, LEF1), and have a polyclonal TCR repertoire, indicating that they behave as stem-like precursors. However, such repertoires are largely distinct, thereby suggesting that these subsets target different specificities. T_{PEX} cells are present in healthy individuals and preferentially develop in response to epitopes from persistent pathogens, indicating that these

long-lived memory T cells adapt to persistent/repeated epitope stimulation by acquiring hybrid characteristics (i.e. simultaneously featuring early memory and dysfunctional/exhausted-like traits). A limitation of the study could be that no history of the blood donor cohort is available apart from age and sex. We have investigated the response to highly prevalent pathogens, like EBV and CMV, and we have shown that T_{PEX} preferentially develop in response to these chronic pathogens. Notably, prevalence of EBV is >90% in the population[179], and is thus thought to influence the cohort only marginally. Differently, prevalence of CMV is 60% ca. in the European region[180]. Future investigations will have to address the impact of CMV infection on the distribution of T_{PEX} and T_{STEM} subsets among $CD8^+$ T cells. Notably, we observed a significantly higher proliferation of T_{PEX} cells compared to T_{EM}, implying that these cells are not terminally differentiated, rather are still capable to expand and exert effector functions in response to their cognate Ag. However, while sorted T_{STEM} cells gave origin to the full spectrum of cells on the basis of PD-1 and TIGIT expression, T_{PEX} cells generated only $PD\text{-}1^+TIGIT^+$ cells suggesting a decreased differentiation plasticity that was further corroborated by transcriptional and epigenetic investigations. These results imply that the epigenetic and gene expression profiles of T_{STEM} cells is more dynamic and plastic compared to T_{PEX} cells, which are instead committed to the generation of dysfunctional cells. Although the benefit of the generation of a dysfunctional lineage might be not clear, T_{PEX} are still capable to exert effector activities, albeit at lower magnitude compared to T_{STEM}. This scenario is also reminiscent of T cell functionality in response to chronic infections, where exhausted T cells, despite limited in their effector functions, are still capable to exert some control of viral replication. In line

with this observation, Jin et al have shown that chronically SIV infected macaques experience increased viral load following $CD8^+$ T cell depletion[181]. More recently, Alfei et al have shown that *Tox*-deleted virus-specific T cells abrogate the exhaustion program, undergo effector differentiation and induce lung tissue destruction[101], suggesting that exhaustion indeed limits immunopathology while ensuring some level of persistence and viral control. Importantly, upon activation, T_{STEM} co-express stem and effector genes while being relatively resistant to exhaustion[178], thus suggesting that the promotion of a stem/effector hybrid phenotype would be beneficial for long-term anti-tumor immunity. This is reminiscent of recent data from Lynn and colleagues, who generated exhaustion-resistant CAR T cells by overexpressing c-Jun[182]. Further studies will be required to test whether induction of the transcriptional program of T_{STEM} may result in similar functional capacity.

Overall our work demonstrates that T_{PEX} cells represent a population with unique hybrid characteristics that localizes within the very early stages of the “exhaustion differentiation branch” and that so far was included within the classical memory pool (i.e. T_{SCM} and T_{CM}). We propose that pre-existing differences on gene expression profile and epigenetic imprinting along with TCR specific for chronic antigens allow the commitment to $CCR7^+$ $GZMK^+$ PD-1^+ $TIGIT^+$ T_{PEX} cells that may represent the initial step of exhaustion differentiation.

6 References

1. Morris, G. P. & Allen, P. M. How the TCR balances sensitivity and specificity for the recognition of self and pathogens. *Nat. Immunol.* **13**, 121–8 (2012).
2. Van Der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nature Reviews Immunology* **11**, 47–55 (2011).
3. Robey, E. & Fowlkes, B. J. Selective Events in T Cell Development. *Annu. Rev. Immunol.* **12**, 675–705 (1994).
4. Scott-Browne, J. P., White, J., Kappler, J. W., Gapin, L. & Marrack, P. Germline-encoded amino acids in the αB T-cell receptor control thymic selection. *Nature* **458**, 1043–1046 (2009).
5. Morris, G. P., Ni, P. P. & Allen, P. M. Alloreactivity is limited by the endogenous peptide repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3695–700 (2011).
6. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).
7. Wooldridge, L. *et al.* A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* **287**, 1168–1177 (2012).
8. Krogsgaard, M., Juang, J. & Davis, M. M. A role for 'self' in T-cell activation. *Seminars in Immunology* **19**, 236–244 (2007).
9. Hogquist, K. A. & Jameson, S. C. The self-obsession of T cells: How TCR signaling thresholds affect fate 'decisions' and effector function. *Nat. Immunol.* **15**, 815–823 (2014).
10. Surh, C. D. & Sprent, J. Homeostasis of Naive and Memory T Cells. *Immunity* **29**, 848–862 (2008).

11. Takada, K. & Jameson, S. C. Naive T cell homeostasis: from awareness of space to a sense of place. *Nat. Rev. Immunol.* **9**, 823–832 (2009).

12. Bouneaud, C., Kourilsky, P. & Bousso, P. Impact of negative selection on the T cell repertoire reactive to a self-peptide: A large fraction of T cell clones escapes clonal deletion. *Immunity* **13**, 829–840 (2000).

13. Pircher, H., Rohrer, U. H., Moskophidis, D., Zinkernagel, R. M. & Hengartner, H. Lower receptor avidity required for thymic clonal deletion than for effector T-cell function. *Nature* **351**, 482–485 (1991).

14. Oldstone, M. B. A., Nerenberg, M., Southern, P., Price, J. & Lewicki, H. Virus infection triggers insulin-dependent diabetes mellitus in a transgenic model: Role of anti-self (virus) immune response. *Cell* **65**, 319–331 (1991).

15. Redmond, W. L. & Sherman, L. A. Peripheral tolerance of CD8 T lymphocytes. *Immunity* **22**, 275–284 (2005).

16. Mannie, M. D. A unified model for T cell antigen recognition and thymic selection of the T cell repertoire. *J. Theor. Biol.* **151**, 169–192 (1991).

17. Harty, J. T. & Badovinac, V. P. Shaping and reshaping CD8+ T-cell memory. *Nature Reviews Immunology* **8**, 107–119 (2008).

18. Sprent, J. & Surh, C. D. T C ELL M EMORY . *Annu. Rev. Immunol.* **20**, 551–579 (2002).

19. Sallusto, F., Geginat, J. & Lanzavecchia, A. C entral M emory and E ffector M emory T C ell S ubsets : Function, Generation, and Maintenance . *Annu. Rev. Immunol.* **22**, 745–763 (2004).

20. Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999).

21. Reinhardt, R. L., Khoruts, A., Merica, R., Zell, T. & Jenkins, M. K. Visualizing the generation of memory CD4 T cells in the whole body. *Nature* **410**, 101–5 (2001).

22. Masopust, D., Vezys, V., Marzo, A. L. & Lefrançois, L. Preferential localization of effector memory cells in nonlymphoid tissue. *J. Immunol.* **192**, 845–849 (2014).

23. Buchholz, V. R., Schumacher, T. N. M. & Busch, D. H. T Cell Fate at the Single-Cell Level. *Annu. Rev. Immunol.* **34**, 65–92 (2016).

24. Gattinoni, L. *et al.* Wnt signaling arrests effector T cell differentiation and generates CD8 + memory stem cells. *Nat. Med.* **15**, 808–813 (2009).

25. Gattinoni, L. *et al.* A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297 (2011).

26. Lugli, E. *et al.* Identification, isolation and in vitro expansion of human and nonhuman primate T stem cell memory cells. *Nat. Protoc.* **8**, 33–42 (2013).

27. Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M. & Lugli, E. The who's who of T-cell differentiation: Human memory T-cell subsets. *Eur. J. Immunol.* **43**, 2797–2809 (2013).

28. Appay, V., van Lier, R. a W., Sallusto, F. & Roederer, M. Phenot1. Appay, V., van Lier, R. a W., Sallusto, F. & Roederer, M. Phenotype and function of human T lymphocyte subsets: consensus and issues. Cytometry. A 73, 975–83 (2008).ype and function of human T lymphocyte subsets: consensus and issues. *Cytometry. A* **73**,

975–83 (2008).

29. Romero, P. *et al.* Four Functionally Distinct Populations of Human Effector-Memory CD8 + T Lymphocytes . *J. Immunol.* **178**, 4112–4119 (2007).

30. Lugli, E. *et al.* Superior T memory stem cell persistence supports long-lived T cell memory. *J. Clin. Invest.* **123**, 594–9 (2013).

31. Jameson, S. C. Maintaining the norm: T-cell homeostasis. *Nat. Rev. Immunol.* **2**, 547–56 (2002).

32. Takeda, S., Rodewald, H. R., Arakawa, H., Bluethmann, H. & Shimizu, T. MHC class II molecules are not required for survival of newly generated CD4+ T cells, but affect their long-term life span. *Immunity* **5**, 217–228 (1996).

33. Tanchot, C., Lemonnier, F. A., Pérarnau, B., Freitas, A. A. & Rocha, B. Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science (80-.).* **276**, 2057–2062 (1997).

34. Tan, J. T. *et al.* IL-7 is critical for homeostatic proliferation and survival of naive T cells. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8732–7 (2001).

35. Vivien, L., Benoist, C. & Mathis, D. T lymphocytes need IL-7 but not IL-4 or IL-6 to survive in vivo. *Int. Immunol.* **13**, 763–8 (2001).

36. Kieper, W. C. *et al.* Overexpression of interleukin (IL)-7 leads to IL-15-independent generation of memory phenotype CD8+ T cells. *J. Exp. Med.* **195**, 1533–1539 (2002).

37. Mazzucchelli, R. & Durum, S. K. Interleukin-7 receptor expression: Intelligent design. *Nature Reviews Immunology* **7**, 144–154 (2007).

38. Khaled, A. R. & Durum, S. K. Lymphocide: Cytokines and the control of lymphoid homeostasis. *Nature Reviews Immunology* **2**, 817–830

(2002).

39. Qin, J. Z. *et al.* Interleukin-7 and interleukin-15 regulate the expression of the bcl-2 and c-Myb genes in cutaneous T-cell lymphoma cells. *Blood* **98**, 2778–2783 (2001).

40. Napolitano, L. A. *et al.* Increased production of IL-7 accompanies HIV-1-mediated T-cell depletion: Implications for T-cell homeostasis. *Nat. Med.* **7**, 73–79 (2001).

41. Fry, T. J. *et al.* A potential role for interleukin-7 in T-cell homeostasis. *Blood* **97**, 2983–2990 (2001).

42. Surh, C. D. & Sprent, J. Regulation of mature T cell homeostasis. *Semin. Immunol.* **17**, 183–191 (2005).

43. Min, B. *et al.* Neonates support lymphopenia-induced proliferation. *Immunity* **18**, 131–140 (2003).

44. Goldrath, A. W. & Bevan, M. J. Low-affinity ligands for the TCR drive proliferation of mature CD8+ T cells in lymphopenic hosts. *Immunity* **11**, 183–190 (1999).

45. Kieper, W. C., Burghardt, J. T. & Surh, C. D. A Role for TCR Affinity in Regulating Naive T Cell Homeostasis. *J. Immunol.* **172**, 40–44 (2004).

46. Workman, C. J. & Vignali, D. A. A. Negative Regulation of T Cell Homeostasis by Lymphocyte Activation Gene-3 (CD223). *J. Immunol.* **174**, 688–695 (2005).

47. Krieg, C., Boyman, O., Fu, Y. X. & Kaye, J. B and T lymphocyte attenuator regulates CD8+ T cell-intrinsic homeostasis and memory cell generation. *Nat. Immunol.* **8**, 162–171 (2007).

48. Posevitz, V. *et al.* Regulation of T Cell Homeostasis by the Transmembrane Adaptor Protein SIT. *J. Immunol.* **180**, 1634–1642

(2008).

49. Li, O., Zheng, P. & Liu, Y. CD24 expression on T cells is required for optimal T cell proliferation in lymphopenic host. *J. Exp. Med.* **200**, 1083–1089 (2004).

50. Ring, A. M. *et al.* Mechanistic and structural insight into the functional dichotomy between IL-2 and IL-15. *Nat. Immunol.* **13**, 1187–95 (2012).

51. Geginat, J., Lanzavecchia, A. & Sallusto, F. Proliferation and differentiation potential of human CD8+ memory T-cell subsets in response to antigen or homeostatic cytokines. *Blood* **101**, 4260–4266 (2003).

52. Ma, A., Koka, R. & Burkett, P. Diverse functions of IL-2, IL-15, and IL-7 in lymphoid homeostasis. *Annu. Rev. Immunol.* **24**, 657–79 (2006).

53. Cho, J. H. *et al.* An intense form of homeostatic proliferation of naive CD8+ cells driven by IL-2. *J. Exp. Med.* **204**, 1787–1801 (2007).

54. Ramsey, C. *et al.* The Lymphopenic Environment of CD132 (Common γ-Chain)-Deficient Hosts Elicits Rapid Homeostatic Proliferation of Naive T Cells via IL-15. *J. Immunol.* **180**, 5320–5326 (2008).

55. Berard, M., Brandt, K., Paus, S. B. & Tough, D. F. IL-15 Promotes the Survival of Naive and Memory Phenotype CD8 + T Cells . *J. Immunol.* **170**, 5018–5026 (2003).

56. Kieper, W. C. *et al.* Cutting Edge: Recent Immune Status Determines the Source of Antigens That Drive Homeostatic T Cell Expansion. *J. Immunol.* **174**, 3158–3163 (2005).

57. Tough, D. F. & Sprent, J. Turnover of naive- and memory-phenotype

T cells. *J. Exp. Med.* **179**, 1127–1135 (1994).

58. Tough, D. F., Borrow, P. & Sprent, J. Induction of bystander T cell proliferation by viruses and type I interferon in vivo. *Science (80-.).* **272**, 1947–1950 (1996).

59. Tough, D. F., Sun, S. & Sprent, J. T cell stimulation in vivo by lipopolysaccharide (LPS). *J. Exp. Med.* **185**, 2089–2094 (1997).

60. Zhang, X., Sun, S., Hwang, I., Tough, D. F. & Sprent, J. Potent and selective stimulation of memory-phenotype CD8+ T cells in vivo by IL-15. *Immunity* **8**, 591–599 (1998).

61. Intlekofer, A. M. *et al.* Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nat. Immunol.* **6**, 1236–1244 (2005).

62. Becker, T. C. *et al.* Interleukin 15 is required for proliferative renewal of virus-specific memory CD8 T cells. *J. Exp. Med.* **195**, 1541–1548 (2002).

63. Lenz, D. C. *et al.* IL-7 regulates basal homeostatic proliferation of antiviral CD4 +T cell memory. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9357–9362 (2004).

64. Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M. & Lugli, E. The who's who of T-cell differentiation: Human memory T-cell subsets. *Eur. J. Immunol.* **43**, 2797–2809 (2013).

65. Kaech, S. M. & Ahmed, R. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells. *Nat. Immunol.* **2**, 415–22 (2001).

66. Van Den Broek, T., Borghans, J. A. M. & Van Wijk, F. The full spectrum of human naive T cells. *Nature Reviews Immunology* **18**, 363–373 (2018).

67. Buchholz, V. R. *et al.* Disparate individual fates compose robust

CD8+ T cell immunity. *Science (80-.).* **340**, 630–635 (2013).

68. Gerlach, C. *et al.* Heterogeneous Differentiation Patterns of Individual CD8 + T Cells. **890**, 635–639 (2013).

69. Tubo, N. J. *et al.* Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* **153**, 785–96 (2013).

70. Mandl, J. N., Monteiro, J. P., Vrisekoop, N. & Germain, R. N. T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. *Immunity* **38**, 263–274 (2013).

71. Persaud, S. P., Parker, C. R., Lo, W.-L., Weber, K. S. & Allen, P. M. Intrinsic CD4+ T cell sensitivity and response to a pathogen are set and sustained by avidity for thymic and peripheral complexes of self peptide and MHC. *Nat. Immunol.* **15**, 266–74 (2014).

72. Fulton, R. B. *et al.* The TCR's sensitivity to self peptide-MHC dictates the ability of naive CD8(+) T cells to respond to foreign antigens. *Nat. Immunol.* **16**, 107–17 (2015).

73. Cho, J. H., Kim, H. O., Surh, C. D. & Sprent, J. T Cell Receptor-Dependent Regulation of Lipid Rafts Controls Naive CD8+ T Cell Homeostasis. *Immunity* **32**, 214–226 (2010).

74. Alanio, C. *et al.* Bystander hyperactivation of preimmune CD8+ T cells in chronic HCV patients. *Elife* **4**, (2015).

75. Blank, C. U. *et al.* Defining 'T cell exhaustion'. *Nat. Rev. Immunol.* **19**, 665–674 (2019).

76. Wherry, E. J. T cell exhaustion. *Nature Immunology* **12**, 492–499 (2011).

77. Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. &

Ahmed, R. Viral Persistence Alters CD8 T-Cell Immunodominance and Tissue Distribution and Results in Distinct Stages of Functional Impairment. *J. Virol.* **77**, 4911–4927 (2003).

78. Frebel, H. *et al.* Programmed death 1 protects from fatal circulatory failure during systemic virus infection of mice. *J. Exp. Med.* **209**, 2485–2499 (2012).
79. Cornberg, M. *et al.* Clonal Exhaustion as a Mechanism to Protect Against Severe Immunopathology and Death from an Overwhelming CD8 T Cell Response. *Front. Immunol.* **4**, 475 (2013).
80. Paley, M. A. *et al.* Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science (80-.).* **338**, 1220–1225 (2012).
81. He, R. *et al.* Follicular CXCR5-expressing CD8+ T cells curtail chronic viral infection. *Nature* **537**, 412–416 (2016).
82. Im, S. J. *et al.* Defining CD8 + T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417–421 (2016).
83. Utzschneider, D. T. *et al.* T Cell Factor 1-Expressing Memory-like CD8(+) T Cells Sustain the Immune Response to Chronic Viral Infections. *Immunity* **45**, 415–27 (2016).
84. Wu, T. *et al.* The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. *Sci. Immunol.* **1**, (2016).
85. Wieland, D. *et al.* TCF1+ hepatitis C virus-specific CD8+ T cells are maintained after cessation of chronic antigen stimulation. *Nat. Commun.* **8**, 1–13 (2017).
86. Leong, Y. A. *et al.* CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. *Nat. Immunol.* **17**, 1187–1196 (2016).

87. Baitsch, L. *et al.* Exhaustion of tumor-specific CD8+ T cells in metastases from melanoma patients. *J. Clin. Invest.* **121**, 2350–2360 (2011).

88. Brummelman, J. *et al.* High-dimensional single cell analysis identifies stem-like cytotoxic CD8+ T cells infiltrating human tumors. *J. Exp. Med.* **215**, 2520–2535 (2018).

89. Sade-Feldman, M. *et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* **175**, 998-1013.e20 (2018).

90. Siddiqui, I. *et al.* Intratumoral Tcf1 + PD-1 + CD8 + T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. *Immunity* **50**, 195-211.e10 (2019).

91. Kurtulus, S. *et al.* Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1 − CD8 + Tumor-Infiltrating T Cells. *Immunity* **50**, 181-194.e6 (2019).

92. Li, H. *et al.* Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell* **176**, 775-789.e18 (2019).

93. Henning, A. N., Roychoudhuri, R. & Restifo, N. P. Epigenetic control of $CD8^+$ T cell differentiation. *Nat Rev Immunol2* **18**, 340–356 (2018).

94. Pauken, K. E. *et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160–1165 (2016).

95. Sen, D. R. *et al.* The epigenetic landscape of T cell exhaustion - suppmemental materials. *Science (80-.).* (2016).

doi:10.1126/science.aae0491

96. Scott-Browne, J. P. *et al.* Dynamic Changes in Chromatin Accessibility Occur in CD8+ T Cells Responding to Viral Infection. *Immunity* **45**, 1327–1340 (2016).

97. Philip, M. *et al.* Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* **545**, 452–456 (2017).

98. Mognol, G. P. *et al.* Exhaustion-associated regulatory regions in CD8+ tumor-infiltrating T cells. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2776–E2785 (2017).

99. Alfei, F. *et al.* TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 265–269 (2019).

100. Khan, O. *et al.* TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* **571**, 211–218 (2019).

101. Scott, A. C. *et al.* TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270–274 (2019).

102. Yao, C. *et al.* Single-cell RNA-seq reveals TOX as a key regulator of CD8+ T cell persistence in chronic infection. *Nat. Immunol.* **20**, 890–901 (2019).

103. Wang, X. *et al.* TOX promotes the exhaustion of antitumor CD8+ T cells by preventing PD1 degradation in hepatocellular carcinoma. *J. Hepatol.* **71**, 731–741 (2019).

104. De Simone, G. *et al.* CXCR3 Identifies Human Naive CD8 + T Cells with Enhanced Effector Differentiation Potential . *J. Immunol.* **203**, 3179–3189 (2019).

105. Galletti, G. *et al.* Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans. *Nat. Immunol.*

21, 1552–1562 (2020).

106. Lugli, E. *et al.* IL-15 delays suppression and fails to promote immune reconstitution in virally suppressed chronically SIV-infected macaques. *Blood* **118**, 2520–2529 (2011).
107. Cossarizza, A. *et al.* Guidelines for the use of flow cytometry and cell sorting in immunological studies. *Eur. J. Immunol.* **47**, 1584–1797 (2017).
108. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2**, e132 (2006).
109. Orrù, V. *et al.* XGenetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242 (2013).
110. Douek, D. C. *et al.* Changes in thymic function with age and during the treatment of HIV infection. *Nature* **396**, 690–5 (1998).
111. Roberto, A. *et al.* Role of naive-derived T memory stem cells in T-cell reconstitution following allogeneic transplantation. *Blood* **125**, 2855–2864 (2015).
112. Geiger, R., Duhen, T., Lanzavecchia, A. & Sallusto, F. Human naive and memory CD4 $^{+}$ T cell repertoires specific for naturally processed antigens analyzed using libraries of amplified T cells. *J. Exp. Med.* **206**, 1525–1534 (2009).
113. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
114. Bolotin, D. A. *et al.* MiXCR: Software for comprehensive adaptive immunity profiling. *Nature Methods* **12**, 380–381 (2015).
115. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
116. Miyazawa, S. & Jernigan, R. L. Residue-residue potentials with a

favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).

117. Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J. Protein Chem.* **4**, 23–55 (1985).
118. Rackovsky, S. Global characteristics of protein sequences and their implications. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8623–8626 (2010).
119. Roberto, A. *et al.* The early expansion of anergic NKG2A pos /CD56 dim /CD16 neg natural killer represents a therapeutic target in haploidentical hematopoietic stem cell transplantation. *Haematologica* **103**, 1390–1402 (2018).
120. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
121. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
122. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
123. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
124. Wong, M. T. *et al.* A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. *Immunity* **45**, 442–456 (2016).
125. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A.

Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

126. Newell, E. W. *et al.* Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat. Biotechnol.* **31**, 623–629 (2013).
127. Bakker, A. H. *et al.* Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3825–3830 (2008).
128. Capper, R. *et al.* The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev.* **21**, 2495–2508 (2007).
129. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
130. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).
131. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).
132. Brodie, T., Brenna, E. & Sallusto, F. OMIP-018: Chemokine receptor expression on human T helper cells. *Cytom. Part A* **83 A**, 530–532 (2013).
133. Thome, J. J. C. *et al.* Spatial map of human t cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814–828 (2014).
134. Linton, P. J. & Dorshkind, K. Age-related changes in lymphocyte development and function. *Nature Immunology* **5**, 133–139 (2004).

135. Haynes, L., Eaton, S. M., Burns, E. M., Rincon, M. & Swain, S. L. Inflammatory Cytokines Overcome Age-Related Defects in CD4 T Cell Responses In Vivo. *J. Immunol.* **172**, 5194–5199 (2004).

136. Song, K. *et al.* Characterization of subsets of CD4+ memory T cells reveals early branched pathways of T cell differentiation in humans. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7916–7921 (2005).

137. Murata, K. *et al.* Identification of a novel human memory T cell population with the characteristics of stem-like chemo-resistance. *Sapporo Medical journal* **86**, 110–111 (2017).

138. McFarland, R. D., Douek, D. C., Koup, R. A. & Picker, L. J. Identification of a human recent thymic emigrant phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4215–4220 (2000).

139. Pittet, M. J. *et al.* High frequencies of naive Melan-A/MART-1-specific CD8+ T cells in a large proportion of human histocompatibility leukocyte antigen (HLA)-A2 individuals. *J. Exp. Med.* **190**, 705–715 (1999).

140. Zippelius, A. *et al.* Thymic selection generates a large T cell pool recognizing a self-peptide in humans. *J. Exp. Med.* **195**, 485–494 (2002).

141. Pulko, V. *et al.* Human memory T cells with a naive phenotype accumulate with aging and respond to persistent viruses. *Nat. Immunol.* **17**, 966–975 (2016).

142. O'Brien, S. *et al.* Ikaros Imposes a Barrier to CD8 + T Cell Differentiation by Restricting Autocrine IL-2 Production . *J. Immunol.* **192**, 5118–5129 (2014).

143. Egorov, E. S. *et al.* The changing landscape of naive T cell receptor repertoire with human aging. *Front. Immunol.* **9**, (2018).

144. Košmrlj, A., Jha, A. K., Huseby, E. S., Kardar, M. & Chakraborty, A. K. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16671–16676 (2008).

145. Feng, Y. *et al.* A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. *Nature* **528**, 132–136 (2015).

146. Izraelson, M. *et al.* Comparative analysis of murine T-cell receptor repertoires. *Immunology* **153**, 133–144 (2018).

147. Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).

148. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–47 (2019).

149. Dusseaux, M. *et al.* Human MAIT cells are xenobiotic-resistant, tissue-targeted, CD161 hi IL-17-secreting T cells. *Blood* **117**, 1250–1259 (2011).

150. Bin Dhuban, K. *et al.* Coexpression of TIGIT and FCRL3 Identifies Helios + Human Memory Regulatory T Cells . *J. Immunol.* **194**, 3687–3696 (2015).

151. Oberdoerffer, S. *et al.* Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science (80-.).* **321**, 686–691 (2008).

152. Stephen, T. L. *et al.* SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. *Immunity* **46**, 51–64 (2017).

153. Giordano, M. *et al.* Molecular profiling of CD 8 T cells in autochthonous melanoma identifies Maf as driver of exhaustion .

EMBO J. **34**, 2042–2058 (2015).

154. Li, J., He, Y., Hao, J., Ni, L. & Dong, C. High Levels of Eomes Promote Exhaustion of Anti-tumor CD8+ T Cells. *Front. Immunol.* **9**, 2981 (2018).

155. Seo, H. *et al.* TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8+ T cell exhaustion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12410–12415 (2019).

156. Kondo, T. *et al.* Notch-mediated conversion of activated T cells into stem cell memory-like T cells for adoptive immunotherapy. *Nat. Commun.* **8**, 1–14 (2017).

157. Widjaja, C. E. *et al.* Proteasome activity regulates CD8+ T lymphocyte metabolism and fate specification. in *Journal of Clinical Investigation* **127**, 3609–3623 (American Society for Clinical Investigation, 2017).

158. Yang, Z. Z. *et al.* TGF-β upregulates CD70 expression and induces exhaustion of effector memory T cells in B-cell non-Hodgkin's lymphoma. *Leukemia* **28**, 1872–1884 (2014).

159. Vodnala, S. K. *et al.* T cell stemness and dysfunction in tumors are triggered by a common mechanism. *Science (80-.).* **363**, (2019).

160. Doering, T. A. *et al.* Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8+ T Cell Exhaustion versus Memory. *Immunity* **37**, 1130–1144 (2012).

161. Utzschneider, D. T. *et al.* High antigen levels induce an exhausted phenotype in a chronic infection without impairing T cell expansion and survival. *J. Exp. Med.* **213**, 1819–1834 (2016).

162. Fuertes Marraco, S. A. *et al.* Long-lasting stem cell-like memory CD8+ T cells with a naïve-like profile upon yellow fever vaccination.

Sci. Transl. Med. **7**, (2015).

163. Akondy, R. S. *et al.* Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).
164. Price, D. A. *et al.* Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. *J. Exp. Med.* **202**, 1349–1361 (2005).
165. Alanio, C. *et al.* CXCR3/CXCL10 Axis Shapes Tissue Distribution of Memory Phenotype CD8+ T Cells in Nonimmunized Mice. *J. Immunol.* **200**, 139–146 (2018).
166. Cose, S., Brammer, C., Khanna, K. M., Masopust, D. & Lefrançois, L. Evidence that a significant number of naive T cells enter non-lymphoid organs as part of a normal migratory pathway. *Eur. J. Immunol.* **36**, 1423–1433 (2006).
167. Akrami, M. *et al.* Circulation of gut-preactivated naïve CD8+T cells enhances antitumor immunity in B cell-defective mice. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23674–23683 (2020).
168. Annunziato, F. *et al.* CXCR3 and αEβ7 integrin identify a subset of CD8+ mature thymocytes that share phenotypic and functional properties with CD8+ gut intraepithelial lymphocytes. *Gut* **55**, 961–968 (2006).
169. Rabin, R. L. *et al.* Chemokine receptor responses on T cells are achieved through regulation of both receptor expression and signaling. *J. Immunol.* **162**, 3840–50 (1999).
170. Groom, J. R. & Luster, A. D. CXCR3 ligands: Redundant, collaborative and antagonistic functions. *Immunology and Cell Biology* **89**, 207–215 (2011).
171. Sung, J. H. *et al.* Chemokine guidance of central memory T cells is

critical for antiviral recall responses in lymph nodes. *Cell* **150**, 1249–63 (2012).

172. Kimmig, S. *et al.* Two subsets of naive T helper cells with distinct T cell receptor excision circle content in human adult peripheral blood. *J. Exp. Med.* **195**, 789–794 (2002).
173. Kato, A., Takaori-Kondo, A., Minato, N. & Hamazaki, Y. CXCR3 high CD8 + T cells with naïve phenotype and high capacity for IFN-γ production are generated during homeostatic T-cell proliferation. *Eur. J. Immunol.* **48**, 1663–1678 (2018).
174. Smith, N. L. *et al.* CD8_naonatal-innate-like_Cell2018. *Cell* **174**, 117-130.e14 (2018).
175. Yu, W. *et al.* Clonal Deletion Prunes but Does Not Eliminate Self-Specific αβ CD8(+) T Lymphocytes. *Immunity* **42**, 929–41 (2015).
176. Gattinoni, L., Klebanoff, C. A. & Restifo, N. P. Paths to stemness: building the ultimate antitumour T cell. *Nat. Rev. Cancer* **12**, 671–84 (2012).
177. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
178. Lugli, E., Galletti, G., Boi, S. K. & Youngblood, B. A. Stem, Effector, and Hybrid States of Memory CD8+ T Cells. *Trends Immunol.* (2019). doi:10.1016/j.it.2019.11.004
179. Smatti, M. K. *et al.* Epstein-barr virus epidemiology, serology, and genetic variability of LMP-1 oncogene among healthy population: An update. *Frontiers in Oncology* **8**, 211 (2018).
180. Zuhair, M. *et al.* Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. (2019).

doi:10.1002/rmv.2034

181. Jin, X. *et al.* Dramatic rise in plasma viremia after CD8+ T cell depletion in simian immunodeficiency virus-infected macaques. *J. Exp. Med.* **189**, 991–998 (1999).

182. Lynn, R. C. *et al.* c-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature* **576**, 293–300 (2019).

7 List of Abbreviations

Ag	Antigen
ACT	Adoptive Cell Transfer
APCs	Antigen-presenting cells
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
BM	Bone marrow
CAR	Chimeric Antigen Receptor
CFSE	5-(and 6)-carboxyfluorescein diacetate succinimidyl ester
CIP	Cytokine-Induced Proliferation
cLIP	Chronic Lymphopenia-Induced Proliferation
CMV	Cytomegalovirus
Ct	Cycle threshold
CTCL	Cutaneous T Cell Lymphoma
CyTOF	Cytometry by Time Of Flight
d	Day
DEG	Differentially Expressed Gene
EDTA	Ethylenediaminetetraacetic Acid
γc	Gamma chain
FDR	False Discovery Rate
Flu	Influenza virus
GSEA	Gene Set Enrichment Analysis
HSC	Haematopoietic Stem Cell
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
LCMV	Lymphocytic Choriomeningitis Virus
LIP	Lymphopenia-Induced Proliferation
LN	Lymph node
M1	Matrix Protein 1
mAb	Monoclonal antibody
MART-1	Melanoma antigen recognized by T cells 1
MFI	Median fluorescence intensity
MHC	Major histocompatibility complex
MAIT	Mucosal-Associated Invariant T cells

MPEC	Memory-Precursor Effector Cell
NL	Naïve-like
PB	Peripheral blood
PBMCs	Peripheral blood mononuclear cells
PMA	Phorbol 12-myristate 13-acetate
p:MHC	Peptide-MHC
RT	Room temperature
SD	Standard Deviation
SEM	Standard error of the mean
SLEC	Short-Lived Effector Cell
SLO	Secondary Lymphoid Organ
T_{CM}	Central memory T cells
TCR	T cell receptor
TGF-β	Transforming Growth Factor-Beta
T_{EFF}	Effector T cells
T_{EM}	Effector memory T cell
T_{EX}	Exhausted T cells
TF	Transcription Factor
TFBM	Transcription Factor Binding Motif
T_{MEM}	Memory T cells
T_{MNP}	Memory T cells with Naïve Phenotype
T_{N}	Naïve T cells
TP_{EX}	Progenitors of Exhausted T cells
TREC	T cell Receptor Excision Circle
T_{REG}	Regulatory T cells
T_{RTE}	Recent thymic emigrant T cells
T_{SCM}	T memory stem cells
t-SNE	t-Distributed Stochastic Neighbor Embedding
T_{STEM}	Stem-like T cells
T_{TE}	Terminal effector memory T cells
T_{TM}	Transitional memory T cells
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier
WT	Wild-Type
YFV	Yellow-Fever Virus

8 Author Contribution

E. Lugli and I conceived the study, analyzed and interpreted the data of the first part of the thesis. G. Galletti and E. Lugli conceived the study of the second part of the thesis. G. Galletti, E. Lugli and I analyzed and interpreted the data of the second part of the thesis. Other members of the Laboratory of Translational Immunology supported the experiments. E.M.C Mazza and S. Puccio performed bioinformatic analysis. I performed FACS sorting, after previous training by E. Lugli, F.S. Colombo and A. Anselmo. C. Peano and J. Cibella performed sequencing experiments and supported RNA libraries generation. G. Galletti and I generated ATACseq libraries, with support of S. Polletti who provided critical expertise. M. Kuka, M. Iannacone, M. Casucci, C. Mezzanotte and B. Camisa performed LCMV and CAR-19 mouse experiments. D.M. Chudakov, A.N. Davydov and M. Metsger performed TCRseq experiments. A. Cassotta and F. Sallusto performed CMV experiments. T.M. Bi and E.W. Newell performed CyTOF experiments. D.A. Price supported critical discussion and manuscripts writing.